

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:51:15 ; Search time 222.76 Seconds
(without alignments)
1530.574 Million cell updates/sec

Title: US-09-700-770-6
Perfect score: 543
Sequence: 1 ccgcgcgtggaggcgagg.....gttaagagcaaaaaaaaaa 543

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	543	100.0	543	21	AAZ98173 Human signal pepti
2	543	100.0	543	21	AAZ29723 Human lung specifi
3	532.6	98.1	562	19	AAV54620 LU105 specific con
4	529	97.4	570	21	AAZ65103 Membrane-bound pro
5	529	97.4	570	22	AAZ44249 Human PRO1245 (UNQ
6	519	95.6	519	19	AAV54621 LU105 polypeptide
7	241.4	44.5	244	19	AAV54617 LU105 specific pol
8	225	41.4	225	19	AAV54618 LU105 specific pol
9	158	29.1	190	19	AAV54616 LU105 specific pol
10	114	21.0	114	19	AAV54619 LU105 specific pol
11	56.4	10.4	263	19	AAV38067 Human lung tissue

12	56.4	10.4	263	20	AAZ99422 LU103 specific pol
13	56.4	10.4	507	19	AAV38069 Human lung tissue
14	56.4	10.4	507	20	AAZ98424 LU103 specific pol
15	56.4	10.4	507	21	AAZ29721 Human lung specifi
16	56.4	10.4	519	19	AAV38070 Human lung tissue
17	56.4	10.4	519	20	AAZ99425 LU103 specific pol
18	56.4	10.4	531	21	AAZ98174 Human signal pepti
19	50.8	9.4	3957	22	AAA09686 HSV-2 immediate ea
20	50	9.2	58857	21	AAZ58471 Nucleotide sequenc
21	49.6	9.1	357	20	AAV86241 EST clone AA246
22	46.4	8.5	21185	21	AAZ63350 Streptomyces globi
23	46.4	8.5	63164	21	AAZ63348 Streptomyces globi
24	45.6	8.4	1134	13	AAQ23525 cDNA encoding huma
25	45.6	8.4	6453	19	AAV20436 Human C-H-ras Onco
26	45.6	8.4	6453	21	AAZ88574 Human Ras gene. H
27	45.6	8.4	6453	21	AAZ60602 Nucleotide sequenc
28	45.6	8.4	6453	22	AAZ75128 Human Ras DNA. Ho
29	45.4	8.4	2721	12	AAQ10212 BamHI J-I fragment
30	45.4	8.4	2721	12	AAQ10543 BamHI J-I fragment
31	45.4	8.4	8438	15	AAQ73500 DNA encoding Pseud
32	45.4	8.4	15672	12	AAQ10613 Rianodin receptor
33	44.4	8.2	1209	21	AAZ87293 S. venezuelae deso
34	44.4	8.2	1942	21	AAZ58638 Human PRO1434 prot
35	44.4	8.2	1942	21	AAZ51262 Human DNA encoding
36	44.4	8.2	12441	21	AAZ87284 S. venezuelae deso
37	44.4	8.2	13613	21	AAZ87319 S. venezuelae deso
38	44.4	8.2	38506	21	AAZ75633 Nucleotide sequenc
39	44.4	8.2	38506	21	AAZ56001 Recombinant cosmid
40	44	8.1	1133	14	AAZ43306 Human apolipoprote
41	44	8.1	6453	21	AAA48653 Human Ha-ras DNA
42	44	8.1	6995	16	AAQ85466 Ras oncogene/LacZ
43	44	8.1	7110	22	AAZ81361 Quorum sensing con
44	44	8.1	21034	19	AAV62154 HSV-2 strain SB5 C
45	44	8.1	26338	19	AAV62134 HSV-2 strain SB5 C

ALIGNMENTS

RESULT 1
AAZ98173
ID AAZ98173 standard; cDNA; 543 BP.
AC AAZ98173;
XX
XX
DT 11-MAY-2000 (first entry)
XX
DE Human signal peptide containing protein HSP-65 cDNA SEQ ID NO:199.
KW Human; signal peptide-containing protein; HSP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; nontropic; neuroprotective; cardiovascular; hepatotropic;
KW antischistosomal; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's disease; ovulatory defect;
KW muscular dystrophy; ss.
XX
OS Homo sapiens.
XX
XX WO200000610-A2.
PN
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US1484.
XX
PR 26-JUN-1998; 98US-0090762.
PR 31-JUL-1998; 98US-0094983.
PR 01-OCT-1998; 98US-0102686.
PR 11-DEC-1998; 98US-0112129.
XX
PA (INCY-) INCYTE PHARM INC.

XX AA265103; 98US-0090431.
AC 24-JUN-1998; 98US-0090435.
XX 24-JUN-1998; 98US-0090444.
DT 24-JUN-1998; 98US-0090445.
XX 24-JUN-1998; 98US-0090461.
DE 24-JUN-1998; 98US-0090472.
XX 24-JUN-1998; 98US-0090535.
XX 24-JUN-1998; 98US-0090538.
KW 24-JUN-1998; 98US-0090540.
KW 24-JUN-1998; 98US-0090557.
XX 24-JUN-1998; 98US-0090676.
OS 25-JUN-1998; 98US-0090678.
XX 25-JUN-1998; 98US-0090688.
XX 25-JUN-1998; 98US-0090690.
XX 25-JUN-1998; 98US-0090691.
XX 25-JUN-1998; 98US-0090694.
XX 25-JUN-1998; 98US-0090695.
XX 25-JUN-1998; 98US-0090696.
XX 26-JUN-1998; 98US-0090862.
XX 26-JUN-1998; 98US-0090863.
XX 01-JUL-1998; 98US-0091358.
XX 01-JUL-1998; 98US-0091360.
XX 01-JUL-1998; 98US-0091544.
XX 02-JUL-1998; 98US-0091478.
XX 02-JUL-1998; 98US-0091486.
XX 02-JUL-1998; 98US-0091519.
XX 02-JUL-1998; 98US-0091626.
XX 02-JUL-1998; 98US-0091628.
XX 02-JUL-1998; 98US-0091633.
XX 02-JUL-1998; 98US-0091646.
XX 02-JUL-1998; 98US-0091673.
XX 07-JUL-1998; 98US-0091978.
XX 07-JUL-1998; 98US-0091982.
XX 09-JUL-1998; 98US-0092182.
XX 10-JUL-1998; 98US-0092472.
XX 20-JUL-1998; 98US-0093339.
XX 30-JUL-1998; 98US-0094651.
XX 04-AUG-1998; 98US-0095282.
XX 04-AUG-1998; 98US-0095285.
XX 04-AUG-1998; 98US-0095301.
XX 04-AUG-1998; 98US-0095302.
XX 04-AUG-1998; 98US-0095318.
XX 04-AUG-1998; 98US-0095321.
XX 04-AUG-1998; 98US-0095325.
XX 10-AUG-1998; 98US-0095916.
XX 10-AUG-1998; 98US-0095929.
XX 11-AUG-1998; 98US-0096012.
XX 11-AUG-1998; 98US-0096143.
XX 11-AUG-1998; 98US-0096146.
XX 12-AUG-1998; 98US-0096329.
XX 17-AUG-1998; 98US-0096757.
XX 17-AUG-1998; 98US-0096766.
XX 17-AUG-1998; 98US-0096768.
XX 17-AUG-1998; 98US-0096773.
XX 17-AUG-1998; 98US-0096791.
XX 17-AUG-1998; 98US-0096867.
XX 17-AUG-1998; 98US-0096891.
XX 17-AUG-1998; 98US-0096894.
XX 17-AUG-1998; 98US-0096895.
XX 17-AUG-1998; 98US-0096897.
XX 18-AUG-1998; 98US-0096949.
XX 18-AUG-1998; 98US-0096950.
XX 18-AUG-1998; 98US-0096959.
XX 18-AUG-1998; 98US-0096960.
XX 18-AUG-1998; 98US-0097022.
XX 19-AUG-1998; 98US-0097141.
XX 20-AUG-1998; 98US-0097218.
XX 24-AUG-1998; 98US-0097661.
XX 26-AUG-1998; 98US-0097951.
XX 26-AUG-1998; 98US-0097952.
XX 26-AUG-1998; 98US-0097954.
XX 26-AUG-1998; 98US-0097955.
XX 26-AUG-1998; 98US-0097971.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 11-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.

Membrane-bound protein PRO1245 encoding cDNA.
Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
pharmaceutical; receptor immunoadhesin; gene mapping; ss.

Homo sapiens.

WC9963088-A2.

09-DEC-1999.

02-JUN-1999; 99WO-US12252.

02-JUN-1998; 98US-0087607.

02-JUN-1998; 98US-0087609.

02-JUN-1998; 98US-0087759.

03-JUN-1998; 98US-0087827.

04-JUN-1998; 98US-0088021.

04-JUN-1998; 98US-0088025.

04-JUN-1998; 98US-0088028.

04-JUN-1998; 98US-0088029.

04-JUN-1998; 98US-0088030.

04-JUN-1998; 98US-0088033.

04-JUN-1998; 98US-0088326.

05-JUN-1998; 98US-0088167.

05-JUN-1998; 98US-0088202.

05-JUN-1998; 98US-0088212.

05-JUN-1998; 98US-0088217.

09-JUN-1998; 98US-0088655.

10-JUN-1998; 98US-0088722.

10-JUN-1998; 98US-0088730.

10-JUN-1998; 98US-0088734.

10-JUN-1998; 98US-0088738.

10-JUN-1998; 98US-0088740.

10-JUN-1998; 98US-0088741.

10-JUN-1998; 98US-0088742.

10-JUN-1998; 98US-0088810.

10-JUN-1998; 98US-0088811.

10-JUN-1998; 98US-0088824.

10-JUN-1998; 98US-0088825.

10-JUN-1998; 98US-0088826.

11-JUN-1998; 98US-0088858.

11-JUN-1998; 98US-0088861.

11-JUN-1998; 98US-0088863.

11-JUN-1998; 98US-0088876.

12-JUN-1998; 98US-0089090.

12-JUN-1998; 98US-0089105.

16-JUN-1998; 98US-0089440.

16-JUN-1998; 98US-0089512.

16-JUN-1998; 98US-0089534.

17-JUN-1998; 98US-0089532.

17-JUN-1998; 98US-0089538.

17-JUN-1998; 98US-0089598.

17-JUN-1998; 98US-0089599.

17-JUN-1998; 98US-0089600.

17-JUN-1998; 98US-0089653.

18-JUN-1998; 98US-0089801.

18-JUN-1998; 98US-0089907.

18-JUN-1998; 98US-0089908.

19-JUN-1998; 98US-0089947.

19-JUN-1998; 98US-0089948.

19-JUN-1998; 98US-0089949.

19-JUN-1998; 98US-0089952.

22-JUN-1998; 98US-0090240.

22-JUN-1998; 98US-0090252.

22-JUN-1998; 98US-0090254.

23-JUN-1998; 98US-0090349.

23-JUN-1998; 98US-0090355.

24-JUN-1998; 98US-0090429.

```
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.
XX
XX (GETH ) GENENTECH INC.
XX
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
XX WPI: 2000-072883/06.
XX
XX P-PSDB; AAY66757.
XX
PT Membrane-bound proteins and related nucleotide sequences -
XX
XX Claim 2; Fig 289; 822pp; English.
XX
XX The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
XX Sequence 570 BP; 129 A; 190 C; 170 G; 81 T; 0 other;
XX
Query Match 97.4%; Score 529; DB 21; Length 570;
Best Local Similarity 100.0%; Pred. No. 2.4e-99;
Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 gcgaggaccgggtataagaagcctctgtgcttgcgcggcgagcgcgaggtcccccgcgc 74
DB 1 gcgaggaccgggtataagaagcctctgtgcttgcgcggcgagcgcgaggtcccccgcgc 60
QY 75 gccccgagccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 134
DB 61 gccccgagccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
QY 135 tgcagctccgctgctcttctttagtgagctgcgcgcgcgcgcgcgcgcgcgcgcgcgc 194
DB 121 tgcagctccgctgctcttctttagtgagctgcgcgcgcgcgcgcgcgcgcgcgcgcgc 180
QY 195 ggcgtgagtcgcgccggagcgccggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 254
DB 181 ggcgtgagtcgcgccggagcgccggcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
QY 255 aaccgcgtgaagctcctgctgagcagcctgggcatcccccgcgcgcgcgcgcgcgcgcgc 314
DB 241 aaccgcgtgaagctcctgctgagcagcctgggcatcccccgcgcgcgcgcgcgcgcgcgc 300
QY 315 tcccagaagtgctggtcgtgagctggtcccgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 374
DB 301 tcccagaagtgctggtcgtgagctggtcccgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
QY 375 gccctgctggggccctgacagtgcttggctgagccgagactggagcactatcacacctgag 434
DB 361 gccctgctggggccctgacagtgcttggctgagccgagactggagcactatcacacctgag 420
QY 435 gacaagacgctgccaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 494
|||||
```

```
Db 421 gacaagacgctgccaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 480
QY 495 ccttccccccgcccctctcaataaacgtggttaagacacacacacacacacacacacac 543
|||||
Db 481 ccttccccccgcccctctcaataaacgtggttaagacacacacacacacacacacacac 529
RESULT 5
AAF44249
ID AAF44249 standard; cDNA; 570 BP.
XX
XX AAF44249;
XX
XX 02-APR-2001 (first entry)
XX
XX Human PRO1245 (UNQ629) nucleotide sequence SEQ ID NO:407.
XX
XX Human; secreted and transmembrane protein; PRO; cytostatic;
KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW diagnostic assay; ss.
XX
XX Homo sapiens.
XX
XX WO2000073454-A1.
XX
XX 07-DEC-2000.
XX
XX 30-MAR-2000; 2000WO-US06439.
XX
XX 02-JUN-1999; 99WO-US12252.
XX 23-JUN-1999; 99US-0141037.
XX 07-JUL-1999; 99US-0143048.
XX 20-JUL-1999; 99US-0144758.
XX 26-JUL-1999; 99US-0145698.
XX 28-JUL-1999; 99US-0146222.
XX 17-AUG-1999; 99US-0149396.
XX 15-SEP-1999; 99WO-US21090.
XX 15-SEP-1999; 99WO-US21547.
XX 08-OCT-1999; 99US-0158663.
XX 30-NOV-1999; 99WO-US28313.
XX 01-DEC-1999; 99WO-US28301.
XX 16-DEC-1999; 99WO-US30095.
XX 20-DEC-1999; 99WO-US30911.
XX 05-JAN-2000; 2000WO-US00219.
XX 06-JAN-2000; 2000WO-US00376.
XX 11-FEB-2000; 2000WO-US03565.
XX 18-FEB-2000; 2000WO-US04341.
XX 22-FEB-2000; 2000WO-US04414.
XX 24-FEB-2000; 2000WO-US04914.
XX 24-FEB-2000; 2000WO-US05004.
XX 02-MAR-2000; 2000WO-US05841.
XX 15-MAR-2000; 2000WO-US06884.
XX 20-MAR-2000; 2000WO-US07377.
XX
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A,
PI Grimaldi CJ, Gurney AL, Kljavin IJ, Napier MA, Pan J,
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX
XX WPI: 2001-032160/04.
XX P-PSDB; AAB65280.
XX
XX PRO polynucleotides used to produce polypeptides used to target
PT bioactive molecules such as toxins, radiolabels or antibodies, to
PT specific cells, to cause targeted cell death -
XX
XX Claim 2; Fig 289; 935pp; English.
XX
XX The present invention describes human secreted and transmembrane PRO
CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
```


Db 241 aaccgctgaagctcgtgctgagcagcctgggcatcccccgtgaaccacctcatagaggc 300
Qy 315 tccagaagtgtgctgagctggtcccccagccgctggggccgtgaagccctgaag 374
Db 301 tccagaagtgtgctgagcgtggtcccccagccgctggggccgtgaagccctgaag 360
Qy 375 gccctgctggggccctgacagctgttggctgagccgagactggagctatcacactgag 434
Db 361 gccctgctggggccctgacagctgttggctgagccgagactggagctatcacactgag 420
Qy 435 gacaagaagctgcccacccgagagctgtaaaaccccgcgagagaccgtccatcc 494
Db 421 gacaagaagctgcccacccgagagctgtaaaaccccgcgagagaccgtccatcc 480
Qy 495 ccttccccggccctctcaataaacgtggttaagaca 533
Db 481 ccttccccggccctctcaataaacgtggttaagaca 519

RESULT 7
AAV54617
ID AAV54617 standard; cDNA; 244 BP.
XX
AC AAV54617;
XX
DT 30-OCT-1998 (first entry)
XX
DE LU105 specific polynucleotide sequence from clone 1327836.
XX
KW LU105; lung disease marker; immunoassay; lung disease; cancer;
KW blood; plasma; serum; ss.
OS Homo sapiens.
XX
PN W09833926-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-US01766.
XX
PR 31-JAN-1997; 97US-0791710.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
PI Gordon J, Granadosen, Hodges SC, Klass MR, Kratochvil JD;
PI Roberts-Rapp L, Russell JC, Stroupe SD;
XX
DR WPI; 1998-437479/37.
XX
PT New nucleic acid for the lung disease marker LU105 - polypeptides,
PT antibodies and genes, used for diagnosis, prevention, treatment of
PT lung disease, specifically cancer
XX
PS Claim 11; Fig 1; 123pp; English.
XX
CC Sequences shown in AAV54616 to AAV54621 represent LU105 specific
CC polynucleotide sequences. These are used in the method of the invention
CC for detecting target LU105 nucleic acid. The method comprises treating a
CC sample with at least one LU105 specific nucleic acid, or its complement
CC which is at least 50 percent identical with the LU105 specific nucleic
CC acid sequences (AAV54616 to AAV54621). LU105 is a lung disease marker.
CC Cells transformed with a recombinant expression system that contains
CC LU105 specific nucleic acid fragments, are used to express recombinant
CC LU105 polypeptides which are used to raise antibodies. The antibodies are
CC used to detect the LU105 antigen, and correspondingly this antigen is
CC used to detect specific antibodies, in usual immunoassays. The LU105
CC polypeptides and nucleic acid sequences are used for diagnosis, staging,
CC monitoring, prognosis, prevention, treatment and determination of
CC susceptibility to, lung disease, specifically cancer. The LU105
CC polypeptides are also used to screen for specific binding agents, useful
CC therapeutically. LU105 is a marker for lung disease (present at high
CC concentration, in altered form or in an unusual body compartment). LU105

CC can be detected in blood, plasma or serum in an inexpensive, non-invasive
CC test.
XX
SQ Sequence 244 BP; 25 A; 97 C; 82 G; 39 T; 1 other;

Query Match 44.5%; Score 241.4; DB 19; Length 244;
Best Local Similarity 99.2%; Pred. No. 6,9e-41;
Matches 242; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 15 gcgagagccgggtataagaagcctgctgacctgcccgggagccgaggttccccgcgc 74
Db 1 gcgagagccgggtataagaagcctgctgacctgcccgggagccgaggttccccgcgc 60

Qy 75 gccccgagccccgcgccataaagctcgcccccctctctgggctctgctgagccctgtcc 134
Db 61 gccccgagccccgcgccataaagctcgcccccctctctgggctctgctgagccctgtcc 120

Qy 135 tgcagctccgctgctgcttctttagtggtgctggcgaagcctgtgcccagcctgtcgt 194
Db 121 tgcagctccgctgctgcttctttagtggtgctggcgaagcctgtgcccagcctgtcgt 180

Qy 195 gcgctggagtcggcgccgagagccgggagcccttgcccaacccctcggcaaccctc 254
Db 181 gcgctggagtcggcgccgagagccgggagcccttgcccaacccctcggcaaccctc 240

Qy 255 aacc 258
Db 241 aacc 244

RESULT 8
AAV54618
ID AAV54618 standard; cDNA; 225 BP.
XX
AC AAV54618;
XX
DT 30-OCT-1998 (first entry)
XX
DE LU105 specific polynucleotide sequence from clone 1605935.
XX
KW LU105; lung disease marker; immunoassay; lung disease; cancer;
KW blood; plasma; serum; ss.
OS Homo sapiens.
XX
PN W09833926-A1.
XX
PD 06-AUG-1998.
XX
PF 30-JAN-1998; 98WO-US01766.
XX
PR 31-JAN-1997; 97US-0791710.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
PI Gordon J, Granadosen, Hodges SC, Klass MR, Kratochvil JD;
PI Roberts-Rapp L, Russell JC, Stroupe SD;
XX
DR WPI; 1998-437479/37.
XX
PT New nucleic acid for the lung disease marker LU105 - polypeptides,
PT antibodies and genes, used for diagnosis, prevention, treatment of
PT lung disease, specifically cancer
XX
PS Claim 11; Fig 1; 123pp; English.
XX
CC Sequences shown in AAV54616 to AAV54621 represent LU105 specific
CC polynucleotide sequences. These are used in the method of the invention
CC for detecting target LU105 nucleic acid. The method comprises treating a
CC sample with at least one LU105 specific nucleic acid, or its complement
CC which is at least 50 percent identical with the LU105 specific nucleic
CC acid sequences (AAV54616 to AAV54621). LU105 is a lung disease marker.
CC Cells transformed with a recombinant expression system that contains
CC LU105 specific nucleic acid fragments, are used to express recombinant
CC LU105 polypeptides which are used to raise antibodies. The antibodies are
CC used to detect the LU105 antigen, and correspondingly this antigen is
CC used to detect specific antibodies, in usual immunoassays. The LU105
CC polypeptides and nucleic acid sequences are used for diagnosis, staging,
CC monitoring, prognosis, prevention, treatment and determination of
CC susceptibility to, lung disease, specifically cancer. The LU105
CC polypeptides are also used to screen for specific binding agents, useful
CC therapeutically. LU105 is a marker for lung disease (present at high
CC concentration, in altered form or in an unusual body compartment). LU105

RESULT	9	
AAV54616		
ID	AAV54616	standard; cDNA; 190 BP.
XX		
AC	AAV54616;	
XX		
DT	30-OCT-1998	(first entry)
XX		
DE	LU105	specific polynucleotide sequence from clone 3353867.
XX		
XX	LU105;	lung disease marker; immunoassay; lung disease; cancer;
KW	blood; plasma;	serum; ss.
XX		
OS	Homo sapiens.	
XX		
PN	W09833926-A1.	
XX		
PD	06-AUG-1998.	
XX		
PF	30-JAN-1998;	98WO-US01766.
XX		
PR	31-JAN-1997;	97US-0791710.
XX		
PA	(ABEO)	ABBOTT LAB.
XX		
PI	Billig-Medel PA,	Cohen M, Colpitts TL, Friedman PN;
PI	Gordon J,	Granadosen, Hodges SC, Klass MR, Kratochvil JD;
PI	Roberts-Rapp L,	Russell JC, Stroupe SD;
XX		
XX		
DR	WPI;	1998-437479/37.
XX		
PT	New nucleic acid	for the lung disease marker LU105 - polypeptides,
PT	antibodies and genes,	used for diagnosis, prevention, treatment of
PT	lung disease,	specifically cancer

RESULT	10	
AAV54619		
ID	AAV54619	standard; cDNA; 114 BP.
XX		
AC	AAV54619;	
XX		
DT	30-OCT-1998	(first entry)
XX		
DE	LU105	specific polynucleotide sequence from clone 811640.
XX		
KW	LU105;	lung disease marker; immunocassay; lung disease; cancer;
KW	blood; plasma;	serum; ss.
XX		
OS	Homo sapiens.	
XX		
PN	WO9833926-A1.	
XX		
PD	06-AUG-1998.	
XX		
PF	30-JAN-1998;	98WO-US01766.
XX		
XX	31-JAN-1997;	97US-0791710.
XX		
PA	(ABBO)	ABBOTT LAB.
XX		
PI	Billing-Medel PA,	Cohen M, Colpitts TL, Friedman PN;
PI	Gordon J, Granadosen,	Hodges SC, Klass MR, Kratochvil JD;
PI	Roberts-Rapp L,	Russell JC, Stroupe SD;
XX		
XX	WPI;	1998-437479/37.

XX New nucleic acid for the lung disease marker LU105 - polypeptides,
 PT antibodies and genes, used for diagnosis, prevention, treatment of
 PT lung disease, specifically cancer
 XX
 PS Claim 11; Fig 1; 123pp; English.
 CC
 CC Sequences shown in AAV54616 to AAV54621 represent LU105 specific
 CC polynucleotide sequences. These are used in the method of the invention
 CC for detecting target LU105 nucleic acid. The method comprises treating a
 CC sample with at least one LU105 specific nucleic acid, or its complement
 CC which is at least 50 percent identical with the LU105 specific nucleic
 CC acid sequences (AAV54616 to AAV54621). LU105 is a lung disease marker.
 CC Cells transformed with a recombinant expression system that contains
 CC LU105 specific nucleic acid fragments, are used to express recombinant
 CC LU105 polypeptides which are used to raise antibodies. The antibodies are
 CC used to detect the LU105 antigen, and correspondingly this antigen is
 CC used to detect specific antibodies, in usual immunoassays. The LU105
 CC polypeptides and nucleic acid sequences are used for diagnosis, staging,
 CC monitoring, prognosis, prevention, treatment and determination of
 CC susceptibility to, lung disease, specifically cancer. The LU105
 CC polypeptides are also used to screen for specific binding agents, useful
 CC therapeutically. LU105 is a marker for lung disease (present at high
 CC concentration, in altered form or in an unusual body compartment). LU105
 CC can be detected in blood, plasma or serum in an inexpensive, non-invasive
 CC test.
 XX
 SQ Sequence 114 BP; 26 A; 44 C; 29 G; 15 T; 0 other;

Query Match 21.0%; Score 114; DB 19; Length 114;
 Best Local Similarity 100.0%; Pred. No. 5.2e-15;
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY - 419 agcatctacactgagacagagcgtgccaccgcgagggctgaaaccccgccg 478
 Db 1 agcatctacactgagacagagcgtgccaccgcgagggctgaaaccccgccg 60
 QY - 479 ggagaccgtccatccctcccgccctctcaataaacgtggttaagagc 532
 Db 61 ggagaccgtccatccctcccgccctctcaataaacgtggttaagagc 114

RESULT 11
 AAV38067
 ID AAV38067 standard; cDNA; 263 BP.
 AC AAV38067;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:2.
 XX
 KW Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9820143-A1.
 XX
 PD 14-MAY-1998.
 XX
 PF 05-NOV-1997; 97WO-US20680.
 XX
 PR 05-NOV-1996; 96US-0744211.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
 PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
 XX
 DR WPI; 1998-286957/25.
 XX
 PT Lung tissue derived polynucleotide LU103 - useful to detect,

PT diagnose, stage, monitor, prognosis, prevent, treat or determine
 PT pre-disposition to lung disease, e.g. lung cancer
 XX
 PS Claim 1; Page 67; 86pp; English.
 CC
 CC The present sequence represents a polynucleotide specific for lung
 CC tissue gene LU103. A method has been developed for detecting the
 CC presence of a target LU103 polynucleotide in a test sample, comprising:
 CC (a) contacting the sample with at least 1 LU103-specific polynucleotide,
 CC and (b) detecting the target LU103 polynucleotide in the test sample,
 CC where the LU103 polynucleotide has at least 50% identity to the 263,
 CC 263, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to
 CC AAV38070. The methods and products of the present invention may be used
 CC to detect, diagnose, stage, monitor, prognosis, prevent, treat or
 CC determine the predisposition diseases and conditions of the lung, e.g.
 CC lung cancer.
 XX
 SQ Sequence 263 BP; 64 A; 57 C; 71 G; 71 T; 0 other;
 Query Match 10.4%; Score 56.4; DB 19; Length 263;
 Best Local Similarity 60.4%; Pred. No. 0.0028;
 Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
 QY 229 ccctggccacccctcgccaccctcaaccgcgtgaaagctctgctgagcagcctgggca 288
 Db 60 ctctggacaacattcttccctttatggtatccattaaagcttcttctgaaacctctggca 119
 QY 289 tccccgtgaaccacctatagaggctcccgagagtgctggtgagctgggtcccccagg 348
 Db 120 ttctgttgagcaccttggtgggggctaaagaaagtgttaaatgagctgggaccagagg 179
 QY 349 ccgtggggggcgtgaaagccctgaagccctgct 382
 Db 180 ctctgaaagctgtgaaagaaactgctggaggcgct 213
 RESULT 12
 AAX99422
 ID AAX99422 standard; DNA; 263 BP.
 XX
 AC AAX99422;
 XX
 DT 19-OCT-1999 (first entry)
 XX
 DE LU103 specific polynucleotide isolated from clone 1235531.
 XX
 KW LU103; tumour; lung cancer; detection; FISH;
 KW Fluorescent in situ hybridisation; ss.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN US593265-A.
 XX
 PD 17-AUG-1999.
 XX
 PF 05-NOV-1997; 97US-0964725.
 XX
 PR 05-NOV-1997; 97US-0964725.
 XX
 PR 05-NOV-1996; 96US-0744211.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
 PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
 XX
 DR WPI; 1999-468402/39.
 XX
 PT Polynucleotides useful for detecting, diagnosing and monitoring
 PT diseases of the lung such as lung cancer
 XX
 PS Claim 1; Column 47; 36pp; English.

XX The 263 base pair sequence of a LU103 specific polynucleotide was
CC derived from clone 1235531.
CC The polynucleotides and methods disclosed in the invention can be
CC useful for detecting, diagnosing, staging, monitoring or predicting
CC diseases and conditions of the lung, such as lung cancer.
CC The polynucleotides may be used to produce probes for use in
CC fluorescent in situ hybridization (FISH) technology to perform
CC chromosomal analysis and identify cancer specific alterations
CC such as deletions.
XX
SQ Sequence 263 BP; 64 A; 57 C; 71 G; 71 T; 0 other;

Query Match 10.4%; Score 56.4; DB 20; Length 263;
Best Local Similarity 60.4%; Pred. No. 0.0028;
Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 229 ccttgcccaacccctcgccacccctcaaccgctgaagctcctgctgagcagcctgggca 288
DB 60 cctggacaacattctccctttatggtatccattaaagcttcttgaaacctgggca 119

QY 289 tcccggtgaacacacctcatagagggtcccaagaagtgtgtggtgagctgggtcccccagg 348
DB 120 ttctgttgagcacctgtggagggtctaaagaaagtgtgtaaatgagctgggaccagagg 179

QY 349 ccgtggggccgctgaagccctgaagccctgct 382
DB 180 cttctgaagctgtgaagaacctgctggaggcgt 213

RESULT 13
AAV38069
ID AAV38069 standard; cDNA; 507 BP.
XX
AC AAV38069;
XX
DT 14-SEP-1998 (first entry)
XX
DE Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:4.
XX
KW Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 83..364
FT /*tag= a
FT /product= "LU103"
XX
XX W09820143-A1.
XX
XX PD 14-MAY-1998.
XX
XX PF 05-NOV-1997; 97WO-US20680.
XX
XX PR 05-NOV-1996; 96US-0744211.
XX
XX PA (ABBO) ABBOTT LAB.
XX
XX PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
XX PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX WPI; 1998-286957/25.
XX P-PSDB; AAW62068.
XX
XX Lung tissue derived polynucleotide LU103 - useful to detect,
XX diagnose, stage, monitor, prognosis, prevent, treat or determine
XX pre-disposition to lung disease, e.g. lung cancer
XX
XX Claim 1; Page 67; 86pp; English.
XX
XX The present sequence represents a polynucleotide specific for lung

CC tissue gene LU103. A method has been developed for detecting the
CC presence of a target LU103 polynucleotide in a test sample, comprising:
CC (a) contacting the sample with at least 1 LU103-specific polynucleotide,
CC and (b) detecting the target LU103 polynucleotide in the test sample,
CC where the LU103 polynucleotide has at least 50% identity to the 269,
CC 263, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to
CC AAV38070. The methods and products of the present invention may be used
CC to detect, diagnose, stage, monitor, prognosis, prevent, treat or
CC determine the predisposition diseases and conditions of the lung, e.g.
CC lung cancer.
XX
SQ Sequence 507 BP; 137 A; 110 C; 116 G; 144 T; 0 other;

Query Match 10.4%; Score 56.4; DB 19; Length 507;
Best Local Similarity 60.4%; Pred. No. 0.0029;
Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 229 ccttgcccaacccctcgccacccctcaaccgctgaagctcctgctgagcagcctgggca 288
DB 195 cttctgaacacattctccctttatggtatccattaaagcttcttgaaacctgggca 254

QY 289 tcccggtgaacacacctcatagagggtcccaagaagtgtgtggtgagctgggtcccccagg 348
DB 255 ttctgttgagcacctgtggagggtctaaagaaagtgtgtaaatgagctgggaccagagg 314

QY 349 ccgtggggccgctgaagccctgaagccctgct 382
DB 315 cttctgaagctgtgaagaacctgctggaggcgt 348

RESULT 14
AAV99424
ID AAV99424 standard; DNA; 507 BP.
XX
AC AAV99424;
XX
DT 19-OCT-1999 (first entry)
XX
DE LU103 specific polynucleotide consensus sequence.
XX
KW LU103; tumour; lung cancer; detection;
KW Fluorescent in situ hybridisation; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 85..364
FT /*tag= a
FT /product= "Sequence AAV28334 amino acid residue"
XX
XX US5939265-A.
XX
XX PD 17-AUG-1999.
XX
XX PF 05-NOV-1997; 97US-0964725.
XX
XX PR 05-NOV-1997; 97US-0964725.
XX PR 05-NOV-1996; 96US-0744211.
XX
XX PA (ABBO) ABBOTT LAB.
XX
XX PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
XX PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX WPI; 1999-468402/39.
XX P-PSDB; AAV28334.
XX
XX Polynucleotides useful for detecting, diagnosing and monitoring
XX diseases of the lung such as lung cancer
XX
XX Claim 1; Column 47-49; 36pp; English.
XX PS

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:47:27 ; Search time 115.94 Seconds
(without alignments)
4024.018 Million cell updates/sec

Title: US-09-700-770-3
Perfect score: 2060
Sequence: 1 cttgagagctctcaataact.....ttccattgagaaaaa 2060

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUT_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2060	100.0	2061	4	US-09-008-271A-16
2	1356.8	65.9	1910	4	US-08-974-691-7
3	1199.2	58.2	1299	1	US-08-723-938-4
4	1199.2	58.2	1299	2	US-09-080-538-4
5	1142	55.4	1353	4	US-08-974-691-1
6	747	36.3	1651	4	US-08-974-691-5
7	316.6	15.4	2038	2	US-08-631-097-7
8	316.6	15.4	2038	4	US-08-810-712-11
9	316.6	15.4	2465	5	PCT-US92-08090-1
10	168.2	8.2	1240	1	US-08-240-372-2
11	168.2	8.2	2073	4	US-09-032-523-6
12	168.2	8.2	2733	2	US-08-846-021A-6
13	167.8	8.1	2732	6	5217891-14
14	151.8	7.4	2875	1	US-08-328-314-1
15	151.8	7.4	2875	1	US-08-731-045-1
16	139.2	6.8	2454	3	US-09-079-415-1
17	139.2	6.5	1615	1	US-08-360-673-5
18	123.6	6.0	2032	1	US-08-088-633-1
19	123.6	6.0	2032	1	US-08-245-750-1
20	123.6	6.0	2032	1	US-08-441-750-1
21	123.6	6.0	2032	2	US-08-441-751-1
22	123.6	6.0	2032	5	PCT-US92-02521-1
23	88.4	4.3	1800	3	US-09-039-773A-1
24	73	3.5	1245	4	US-09-318-443-1
25	69.4	3.4	1682	4	US-09-318-443-7
26	68.2	3.3	43676	3	US-09-356-952-12
27	67.8	3.3	1536	4	US-09-318-443-5

28	62.8	3.0	1867	2	US-08-607-509-3	Sequence 3, Appli
29	62.8	3.0	1867	2	US-08-634-642-3	Sequence 3, Appli
30	62.8	3.0	1867	3	US-08-989-370-3	Sequence 3, Appli
31	61.8	3.0	2934	3	US-09-149-934-2	Sequence 2, Appli
32	59	2.9	1254	5	PCT-US96-05320A-894	Sequence 894, App
33	58.2	2.8	2365	4	US-09-183-706-42	Sequence 42, Appl
34	58.2	2.8	2365	4	US-09-567-995-42	Sequence 42, Appl
35	58	2.8	1618	2	US-08-533-669A-9	Sequence 9, Appli
36	58	2.8	1618	2	US-08-607-509-1	Sequence 1, Appli
37	58	2.8	1618	2	US-08-454-036-1	Sequence 1, Appli
38	58	2.8	1618	2	US-08-634-642-1	Sequence 1, Appli
39	58	2.8	1618	3	US-08-989-370-1	Sequence 1, Appli
40	58	2.8	1618	5	PCT-US95-05064-1	Sequence 1, Appli
41	57.2	2.8	1374	3	US-08-929-738-1	Sequence 1, Appli
42	57.2	2.8	1614	3	US-08-929-738-2	Sequence 2, Appli
43	56.8	2.8	3825	4	US-09-208-742-3	Sequence 3, Appli
44	55.6	2.7	3408	3	US-09-058-489-14	Sequence 14, Appli
45	55.6	2.7	5322	3	US-09-058-489-13	Sequence 13, Appli

ALIGNMENTS

RESULT 1
US-09-008-271A-16
Sequence 16, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Hillman, Jennifer L.
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2061 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGAST01
CLONE: 877617
SEQUENCE DESCRIPTION: SEQ ID NO: 16 :
US-09-008-271A-16

[illegible]

Qy	1021	tccgctccaggagatgccaactcttcctagtgagcctgctggtttacacacacgatttgatc	1080
Db	1021	TCCCGTCCAGGAGATGCCACTTCTTCAGTGTGCCCTGCTGGTTACACACCGATTGTGATC	1080
Qy	1081	ccaaagctctagctccttcctcagagccaaaggaccaaagtttgccattcaatggaactg	1140
Db	1081	CCAAAGCCTCTAGTCTCTTCAGGCCAATGGACCAAGTTTGCCATTCAATATGGAACGTG	1140
Qy	1141	ggcgggtagatggaatcctgtagcgaggacaaagtactattgtagaatcaagggtggaat	1200
Db	1141	GGCGGTAGATGGAATCCTGAGCGAGGACAAAGCTGACTATTGTGTGGAATCAAGGGTGCAT	1200
Qy	1201	caagtgaatttcggggaggctctctggagcccaagcctggtcttcgcctttgcccattttg	1260
Db	1201	CAGTGATTTGGGGAGGCTCTCTGGAGGCCAGCCGTGCTTCGCTTTTGGCCATTTTG	1260
Qy	1261	algggatatgggctcggttttcccattctgtctgtggaaaggagtctgcgcccgcgatgg	1320
Db	1261	ATGGGATATGGGCCTCGGTTTCCCATCTGTCTGTGGAAGGAGTTCGSCGCCCGATGG	1320
Qy	1321	atgtactggtggagcagggggtatgataagcctgtctctctcttacttaccctaaacagg	1380
Db	1321	ATGTACTGTGGAGCAGGGGCTATTGGATAGCCTGTCTCTCTTTACTTCAACAGGG	1380
Qy	1381	accctgaagacctgatgagagagagactggtcctggggggtcgtcgaccgcgcacactaca	1440
Db	1381	ACCCGTGAAGGCTGATGGAGGAGAGCTGTCTGGGGGGCTCGGACCCGGCACACTACA	1440
Qy	1441	tcccaccctcacctctgtgcagtcacaggtccctgcctactgctgcagatccacatggagc	1500
Db	1441	TCCCACCCTCACCTTCGTGCGAGTCACGGTCCCTGCTACTTGCAGATCCATATGGAGC	1500
Qy	1501	gttgaagtgggcccaaggctgactctctgtgccaaggactgctgcacactcctggata	1560
Db	1501	GTGTGAAGTGGGCCCGCAGGGCTGACTCTCTGTGCAAGGGCTGTCTGCCCATCTCGATA	1560
Qy	1561	cgggcacgtcccctcatcacaggaccactgaggagatccgggcccgtctgacgcagccattg	1620
Db	1561	CGGGCAGCTCCCTCATCACAGGACCCTGAGGAGATCCGGGGCCTGCATGTCAAGCCATTG	1620
Qy	1621	ggggaaatccccctgtgctgggggaagtacatccctgctgctcgaaatccccaaagctcc	1680
Db	1621	GGGGAAATCCCCCTGTGTGGCTGGGGAGTACATCTCTGTCTGGAAATCCCAAAGCTCC	1680
Qy	1681	ccgcagctcctctcctcttctgggggggtctggtttaaacctcaagcccactgattacgtca	1740
Db	1681	CCGCAGTCTCCTTCTTGTGGGGGGTCTGGTTAACTTCACGGCCCATGATTACGTCA	1740
Qy	1741	tccagactactcgaaatggcgctccgcctctctgtctgcggtttccaggccctggatgctcc	1800
Db	1741	TCCAGACTACTCGAAATGGCGTCCGCCCTCTGTTGTCCGGTTTCCAGGCCCTGATGTCC	1800
Qy	1801	ctcgcgtcgaggcccttcttgatcctcgtgtagctctcttggggacgtatgtggccg	1860
Db	1801	CTCGCCTTCGAGGGCCCTTCTTGGATCTCTGGTGACGTCTTCTTGGGACATGATGTGCCG	1860
Qy	1861	tcttcagccgcggggacatgaagacagcgcccgggtgggctgcgcgctgcgactc	1920
Db	1861	TC TTCAGCCGCGGGACATGAAGACAGCGCCCGGGTGGGCTTGGCGCGCTCGCACTC	1920
Qy	1921	gcggagcggacctcgtatggggagagactgcgaggcgagcttcccggggtgagcccaa	1980
Db	1921	GCGGAGCGGACCTTCGATGGGAGAGACTGCGCAGGGCGCACTTCCCGGGTGAGCCCAA	1980
Qy	1981	gtaaaagcatcgagcgggtggctcgcgaggtcctgtaccacagtaaaaaaccattat	2040
Db	1981	GTGAAGCGCATGCGACGGGTGCTCGCGGAGGTCTTCTGTACCCAGATAAAATCCCATAT	2040
Qy	2041	ttccattgaaaaaiaaaaaa 2060	
Db	2041	TTCCATTGAAAAIAAAAAA 2060	

NAME: Cohen, Herbert
REGISTRATION NUMBER: 25,109
REFERENCE/DOCKET NUMBER: 0744.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)463-7700
TELEFAX: (202)473-6915
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: No. 5968816 applicable
ORIGINAL SOURCE:
ORGANISM: homo sapiens
STRAIN: not applicable
INDIVIDUAL ISOLATE: not applicable
DEVELOPMENTAL STAGE: not applicable
HAPOTYPE: not applicable
TISSUE TYPE: blood
CELL TYPE: Luecocyte
CELL LINE: HeLa
ORGANELLE: not applicable
IMMEDIATE SOURCE:
LIBRARY: not applicable
CLONE: not applicable
POSITION IN GENOME:
CHROMOSOME/SEGMENT: not applicable
MAP POSITION: not applicable
UNITS: not applicable
FEATURE:
NAME/KEY: This is the DNA sequence
NAME/KEY: claimed in 15(v1) as the (C
LOCATION: not available
IDENTIFICATION METHOD: experiment-
IDENTIFICATION METHOD: In specificat
OTHER INFORMATION: prevention of IFR
OTHER INFORMATION: promoted cell dea
PUBLICATION INFORMATION: not available
US-08-631-097-7

[illegible]

RESULT 9
 PCT-US92-08090-1
 ; Sequence 1, Application PC/TUS9208090
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, J. N.
 ; TITLE OF INVENTION: Fusion Protein Genes for Treatment of
 ; TITLE OF INVENTION: AIDS
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Kilpatrick & Cody
 ; STREET: 100 Peachtree Street
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: U.S.
 ; ZIP: 30303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US92/08090
 ; FILING DATE: 19920922
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fdbst, Patrea L.
 ; REGISTRATION NUMBER: 31,284
 ; REFERENCE/DOCKET NUMBER: OMR129
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 404-572-6508
 ; TELEFAX: 404-572-6555
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2465 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapien
 ; TISSUE TYPE: Epithelial
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 1..6
 ; OTHER INFORMATION: /note= "Restriction site"
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 2460..2465
 ; OTHER INFORMATION: /note= "Restriction site"
 PCT-US92-08090-1

Query Match 15.4%; Score 316.6; DB 5; Length 2465;
 Best Local Similarity 57.9%; Pred. No. 7.1e-80;
 Matches 602; Conservative 0; Mismatches 404; Indels 33; Gaps 1;
 QY 878 gggggcccccctccctgggagacagccatctctgtacctctcgaactacagggatgt 937
 Db 1391 gggcggtccagccgtgaccagggggccattcccgagggtctcaagaactacatgcacgc 1450
 QY 938 gcagttattgggaaattgggtggagacgctccacaaaactcactgttgccttga 997
 Db 1451 ccagttactacggagattggcattgcggagcgcggcccccagtggttcacagtcgtcttcga 1510
 QY 998 cactggctctccaaatctctgtgggtccgctccagagagatgccacttttcagttgacctg 1057
 Db 1511 caccgggtctctccaaactgtgggtggtccctccatccactgcacactcctgcacatcccttg 1570
 QY 1058 ctggttacacaccagcattgatcccaaaagcctctagctctctccagggccaatgggaccaa 1117

Db 1571 CTGGATCCACACAGTACACAGCGCAAGTCCAGCACCTACGTGAGAAATGGTACCTC 1630
 QY 1118 gtttgcaattcaatgatggaactggcggttagatgggaatcctgagcagaggaagcagctga- 1176
 Db 1631 GTTTTGACATCCACTATGGCTCGGCGAGCCTCTCCGGGTACCTGAGCCAGGACACTGTGTGTC 1690
 QY 1177 -----ctattgtggaatcaagggtgcactcagt 1204
 Db 1691 GGTCCCTGCCAGTCAGCGTCTGAGCCTCTGCGCTGGCGGTGTCAAAGTGGAGAGGCA 1750
 QY 1205 gattttggggagagctctctgggagccagcctggtcttctgctcttttgcctattttagtg 1264
 Db 1751 GGTCTTTGGGAGGCCACCAAGCAGCAGCAGCATCCTTTCATCCAGCCCAAGTTTCGATGG 1810
 QY 1265 gatatgggctcggttttccattctgtgtggaaggagttcgcccccgatggatgt 1324
 Db 1811 CATCTGGGATGGCTTACCCCGCATCTCGTCAACAACGCTGTGCCCGTCTTCGACAA 1870
 QY 1325 actggtgagcaggggctattgataagcctgtcttctctcttttaacctcaacagggagcc 1384
 Db 1871 CCTGATGCAGCAGAAAGCTGGTGGACCAAGACATCTTCTCTTCTACCTGAGCAGGACCC 1930
 QY 1385 tgaagagcctgatggagagagctggtctctggggggtcgaccgcgcacactacatccc 1444
 Db 1931 AGATGCGCAGCCTGGGGGTGAGCTGATGCTGGGTGGCAGACAGACTCCAAGTATTACAGGG 1990
 QY 1445 accctcaacttggtccagtcaggtccctgactgagcagatcccaatggagcgtgt 1504
 Db 1991 TTCTCTGCTTACTGTAATGTACCCGCAAGGCTTACTGGCAGGTCCACCTGGACCAAGT 2050
 QY 1505 gaaggtggggccagggctgactctctgtgccaaaggctgtgctgccatcctctgatacggg 1564
 Db 2051 GGAGGTGGCCAGCGGCTGACCTGTGCAAGAGAGGCTGTGAGGCGCATTTGTGGACACAGG 2110
 QY 1565 cagtcctctcatcacagggaccactgagagatccggggcctgcatgcagccattggggg 1624
 Db 2111 CACTTCCCTCATGTTGGGCCCGGTGATGAGGTGCGCAGAGTGCAGAAGGCCATCGGGGC 2170
 QY 1625 aatcccttctgtggtggggtacatcctgtgtcctgctcggaatacccagaagctccccgc 1684
 Db 2171 CGTGCCGCTGATTCAGGGGAGTACATGATCCCTGTGAGAAGGTTCACCCCTGCCCCG 2230
 QY 1685 agtctctctcttcttgggggggtctggttttaacctcacgccccatgattacgtcatcca 1744
 Db 2231 GATCACACTGAAGCTGGAGGCAAGGCTACAAGCTGTCCCCAGAGGACTACACGCTCAA 2290
 QY 1745 gactactcgaatggcgtccgctctgtgttcgggtttccaggccctggatgtccctcc 1804
 Db 2291 GGTGTCGACAGCGCGGGAAGACCTCTGCTGAGCGGCTTCATCGGCATGGACATCCCGCC 2350
 QY 1805 gcttgcaaggcccttctgtgactcctggtgacgtcttcttggggagcgtatgtggcgcttt 1864
 Db 2351 ACCCAGCGGCGCACCTCTGGATCCTGGGCGACGCTTTCATCGGCGCGCTACTACACTGTGTT 2410
 QY 1865 cgaccgcgggggacatgaag 1883
 Db 2411 TGACCGTGACAACAACAGG 2429

RESULT 10
 US-08-240-372-2
 ; Sequence 2, Application US/08240372
 ; Patent No. 5741665
 ; GENERAL INFORMATION:
 ; APPLICANT: KATO, ELIE K.
 ; APPLICANT: STUART, W. DORSEY
 ; TITLE OF INVENTION: LIGHT-REGULATED PROMOTERS FOR PRODUCTION
 ; TITLE OF INVENTION: OF HETEROLOGOUS PROTEINS IN FILAMENTOUS FUNGI
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORRISON & FOERSTER
 ; STREET: 2000 Pennsylvania Ave. N.W.

Db 200 CAGTTCACCGAGTCTGCTCAATGGACACGAGTCCCAAGAACCCCTCATCAACTACTT 259
QY 932 ggaatgcagatatttttgggaaattggctgggaacgcctccacaaaacttcactgttgc 991
Db 260 GGATATGAATACTTGGCAGTATCTCATTTGGCTCCCAACACAGAACTTCACTGTGAT 319
QY 992 ctttgacactggctcctccactctctgtgggtcccgctccaggagatgccaactttctcagt 1051
Db 320 CTTGCACACTGGCTCTCAACCTCTGGGTCCCTCTGTG-----TACTGCACCTAGGCC 373
QY 1052 gccctgctgtttacacacccatttgatcccaagcctctagctccttccaggccaatgg 1111
Db 374 AGCTGCAAGAGCGACAGAGTTCCAGCCTTCCAGTCCAGCACATACAGCCAGCCAGG 433
QY 1112 gaccaagtttgcattcaatatggaactgggctgggttagatggaatcctgagcgaggacaa 1171
Db 434 TCAATCTTTCTCCATTCAGTAGGAACGGGAGCTTGGCGGATCATTTGAGCGGACCA 493
QY 1172 gctgactattggtggaatcaagggtgcatcagtgattttcggggaggctctctggagcc 1231
Db 494 AGTCTCTGTGAAGGAGTAACCGTGGTGGCCAGCAGTCTTGGAGAAAGTGTACACAGAGCC 553
QY 1232 cagcctgctctcgtcttttggccattttgagtgatggaatattggcctcggtttccattct 1291
Db 554 AGCCAGACCTTTGTGGATGAGAGTTGTGGAATCTGGGCTGGGATACCCCTCCTT 613
QY 1292 gctgtggaagagttggccggcccgatggatgtactggtgagcaggggcttattggataa 1351
Db 614 GGCTGTGGGAGGAGTACTCCAGTATTGTACAACATGATGCTCAGAACCTGTGGGACTT 673
QY 1352 gctgtctctcttttacctcaacagggaccctggaagaccctgatgagagagctggt 1411
Db 674 GCCGATGTTTCTGTACATGAGCAGTAACCCAGAAAGGTGGTCCGGGAGCGAGCTGAT 733
QY 1412 cctgggggctggaccggcagactacatcccaaccctcacctctgctgcagtcacggt 1471
Db 734 TTTTGGAGGCTTACGACCACCTCCATTTCTCTGGGAGCTGAATGGTGGTCCAGTCAACAA 793
QY 1472 cctgctactggcagatccacatgga 1498
Db 794 GCAAGCTTACTGGCAGATTGCACTGGA 820

RESULT 12

US-08-846-021A-6

; Sequence 6, Application US/08846021A

; Patent No. 5948682

; GENERAL INFORMATION:

; APPLICANT: Moloney, Maurice M.

; TITLE OF INVENTION: Preparation of Heterologous Proteins on

; TITLE OF INVENTION: Oil Bodies

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BERESKIN & PARR

; STREET: 40 King Street West

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5H 3Y2

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/846,021A

; FILING DATE: April 25, 1997

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Gravelle, Micheline

; REGISTRATION NUMBER: 40,261

; REFERENCE/DOCKET NUMBER: 9369-039

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 364-7311

; TELEFAX: (416) 361-1398

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2733 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 850..1206

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1444..2729

; US-08-846-021A-6

Query Match

8.2%; Score 168.2; DB 2; Length 2733;

Best Local Similarity 53.0%; Pred. No. 8.6e-38;

Matches 415; Conservative 0; Mismatches 353; Indels 15; Gaps 2;

QY 911 cgtacctctctcgaactacagggatgtgcagtatatttggggaatttgggctgggaacgcc 970
Db 1770 CGTGCCTCTGACCAACTTACCTGGATAGTCACTACTTTGGGAAGATCTACCTCGGACCCC 1829
QY 971 tcacaaaaactcactcgttgcctttgacactggctcctccaactctctgggtcccgccag 1030
Db 1830 GCCCAGGAGTTCACCGTGTGTTGACACTGGCTCTCTGTGACTTCTGGGTACCCCTTA- 1888
QY 1031 gatatgccactcttccagtgtgccctgtcgtttacacccaccgatttgatcccaaacctc 1090
Db 1889 -----TCTACTGCAAGAGCAATGCCGTGCAAAAACACACCGCTTCGACCCGAGAAGTC 1943
QY 1091 tagctctccacggccaatgggaccaagtgttgcatttcaatatgggaactgggcggtaga 1150
Db 1944 GTCACACTTCCAGAACCTGGGCAAGCCCTGTCTATCCACTACGGGACAGGAGCATGCA 2003
QY 1151 tggaaatctgacgagacaaagctgactatattggtggaatcaaggtgcatcagtgatttt 1210
Db 2004 GGGCATCTCTGGGCTATGACACCTGCTACTGTCTCAACATTTGGGACATCCAGCAGACAT 2063
QY 1211 cggggaggctctctgggagagccagcctgctcttgccttttgcctatttgatgggatatt 1270
Db 2064 AGCCCTGAGCACCCAGGAGCGGGGACGCTCTACCTATGCGGATTCGACGGGATCCT 2123
QY 1271 gggcctcggttttccattctgtctgtggaaggagttcgcccccccgatggatgactggt 1330
Db 2124 GGGGATGGCTTACCCCTCGCTCGCTCAGAGTACTCGATATACCCCTGTTTGTGACAACATGAT 2183
QY 1331 ggagcaggggcttattggataaagcctgtcttctccttttacctcaacagggaccctgaaga 1390
Db 2184 GAACAGGCACCTGGTGGCCCAAGACCTGTTCTCGGTTTACATGGACAGGAATGGCCAGGA 2243
QY 1391 gctgatgagagagagctggtcctggggggtcggaccgggcacacatacatccccccct 1450
Db 2244 G-----AGCATGCTCAGCTGGGGGCCATGACCCGCTCTACTACACAGGGTCCCT 2294
QY 1451 caccctctgctcagtcacggtccctgcctactggcagatccacatggagcgtgtgaagt 1510
Db 2295 GCACCTGGGTGCCCGTGACAGTACTGGCAGTTTCACTGTGGACAGTGTCCCAT 2354
QY 1511 gggccagggctgactctctgtgccaagggtgtgctgcctcctcctggatcagcgacgttc 1570
Db 2355 CAGCGGTGTGGTGTGGCTGTGAGGGTGGCTGTGAGGCCATCTTTGGACACGGGACCTC 2414
QY 1571 cctcatcacaggaccactgagagatcccgccctgcatgcagccattggggggaatccc 1630
Db 2415 CAAGCTGTGCGGGCCAGCAGCAGCATCTCAACATCCAGCAGGCATTTGGAGCCACACA 2474
QY 1631 cttgctgctgggagtagatcatcatcctgctcgtcgaaatcccaagctcccccgagcttc 1690

```
Db 2475 GAACAGTACGGTGAGTTGACATCGACTGCGACAACTGAGCTACATGCCACACTGTGTGT 2534
QY 1691 ctt 1693
Db 2535 CTT 2537

RESULT 13
5217891-14
;PATENT NO. 5217891
;APPLICANT: BRAKE, ANTHONY J.;VAN DEN BERG, JOHAN A.
;TITLE OF INVENTION: DNA CONSTRUCTS CONTAINING A KLUYVEROMYCES
;A FACTOR LEADER SEQUENCE FOR DIRECTING SECRETION OF HETEROLOGOUS
;POLYPEPTIDES
;NUMBER OF SEQUENCES: 23
;CURRENT APPLICATION DATA:
;APPLICATION NUMBER: US/08/507,398
;FILING DATE: 09-APR-1990
;PRIOR APPLICATION DATA:
;APPLICATION NUMBER: 78,551
;FILING DATE: 28-JUL-1987
;SEQ ID NO:14:
;LENGTH: 2732
5217891-14

Query Match      8.1%; Score 167.8; DB 6; Length 2732;
Best Local Similarity 52.0%; Pred. No. 1.le-37;
Matches 434; Conservative 0; Mismatches 387; Indels 14; Gaps 2;

QY 911 ggtactctctgaactacaggatgtgcagtatttggggaattgggctgggaacgcc 970
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 828 ggtccctcgaccactacctggacagtcagtcacttgggaagatctacctcggaaccc 887
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 971 tccacaaactcactgttgcctttgacactggtcctcccaatctctgtgggtcccgccag 1030
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 888 gccacagagttcacctgctgtttgacactggtcctctgactctgtgggtaccctcta- 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1031 gagatgccactcttcagtgctccctgctgtgttacacacacgattgtatcccaaaccttc 1090
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 947 -----tctactcaagagcaatgcctgcgaacacacacacgccttcgaccgcgagaaagtc 1001
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1091 tagtctcttcaggccaaaggacacaaagtttgccattcaatatgaactggcggttaga 1150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1002 gtccacctccagaacctgggcaagccctgtctatccactacgagcagcagcatgca 1061
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1151 tgaatactgacgaggaacagcactactattgttggaatcaagggtgcatcagtgatttt 1210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1062 gggcatcctgggctatgacacgcctgactgtctccacattgtggacatccacgacagtc 1121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1211 cggggaggctctctgggagccagcctggtctctctttgcccattttgatgggatatt 1270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1122 aggcctgagcaccagagccggagcgtctctaccatgcggaattcgacgggatcct 1181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1271 gggcctcgttttccattctgtctgtggaagagttggtggcccgatggatgtactggt 1330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1182 ggggatggcctaccctcgctgcctcagagctactgcataaccgctgtttgacacatgat 1241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1331 ggagcaggggctattggataaacctgtctctcttttacctcaacagggacccttgaaga 1390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1242 gaacaggccactgggtggcccaagacctgtctcgttttacctatggacagg-----aat 1293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1391 gcctgatggagagagctggtcctgggggctcgaccgggcacactatcccccct 1450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1294 ggcacaaggagacatgctcaacgctgggggccaatgcaccgctcctactacaaagggtccct 1353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1451 caccctcgtgcagtcacaggtccctgcctactgtgcagatccacatggagcgtgtgaaggt 1510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1354 gcatgggtggccgtgacagtcagcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1511 gggccagggctgactctctgtgcacaagggtgtgtgctgccatcctcgatatacgggcagctc 1570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1414 cagcgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 1571 cctcatcacagaccactgaggagatccgggcccctgcatgcagccattgggggaatccc 1630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1474 caagctggtcgggcccgagcagcgacatcctcaacatccagcaggccattggagccacaca 1533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1631 cttgctggtcggggagtacatcctctgtctcggaatcccaagctccccgcagctctc 1690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1534 gaaccagctacggtgagtttgacatcgactgcgacaacctgagctacatgcccactgtggt 1593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1691 cttcctcttgggggggtctgttttaacctcaacggcccaatgattacgtcatccag 1745
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1594 ctttgatcatcaatgggcaaatgtaccactgaccctccctccctataaccagccag 1648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-08-328-314-1
; Sequence 1, Application US/08328314
; Patent No. 5674728
; GENERAL INFORMATION:
; APPLICANT: Buxton, Frank
; APPLICANT: Jarai, Gabor
; APPLICANT: Visser, Jacob
; TITLE OF INVENTION: Fungal Protease
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,314
; FILING DATE: TBA
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprulli, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 4-19746/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8615
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2875 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: pepe
; STRAIN: Aspergillus niger N400
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1269..1370, 1462..1612, 1669..2323, 2382
; LOCATION: ..2667)
; OTHER INFORMATION: /function= "Aspartic Protease"
; OTHER INFORMATION: /product= "PEPE"
; OTHER INFORMATION: /gene= "pepe"
; FEATURE:
; NAME/KEY: intron
; LOCATION: order(1371..1461, 1613..1668, 2324..2381)
; FEATURE:
; NAME/KEY: exon
; LOCATION: join(1269..1370, 1462..1612, 1669..2323, 2382
; LOCATION: ..2667)
```



```

Db 2023 AGATTGTTCTCCCTTCTACAAACATGCTTGACAGGGGCTCCTCGACGAGCGGTCCTTG 2082
QY 1363 ccttttacctcaacaggaccctgaagacctgatggaggagagctgtcctgggggct 1422
Db 2083 CCTTTACCTTGGAGACACCAACAAGGAGGGTGACGAGTCCGTGGGACCTTCGGTGGTG 2142
QY 1423 cggaccgggcacactacatccccaccctcaccttcgtgccagtcacggtccctgcctact 1482
Db 2143 TCGACACAGGACCCTACACCGGCGAGCTGATCAAGATCCCCCTCCGGCGCAAGGCTTACT 2202
QY 1483 ggcagatcoacatggagcgtgtgaagggtggcccgaggctgactctctgtgccaaaggct 1542
Db 2203 GGGAGGTTGAGCTTGACGCCATTGCTCTTGGCGGATGATTTGCTGAGATGG---AGAAC 2259
QY 1543 gtgtgccatcctgatacgggcacgctccctcactcacaggaccactga 1591
Db 2260 CCGGTGTCATTCTGGACACTGGTACCTCCCTGATTGCTCTGCCTGCTGA 2308
```

Search completed: November 17, 2001, 14:48:23
Job time: 7181 sec

THIS PAGE BLANK (uspro)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 12:48:42 ; Search time 115.94 Seconds
(without alignments)
990.377 Million cell updates/sec

Title: US-09-700-770-1
Perfect score: 507
Sequence: 1 ggcaagtggaccactggct.....aataaagcaatgaatacatt 507

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	507	100.0	507	2	US-08-964-725-4
2	499.4	98.5	519	2	US-08-964-725-5
3	263	51.9	263	2	US-08-964-725-2
4	263	51.9	269	2	US-08-964-725-1
5	225	44.4	225	2	US-08-964-725-3
6	38	7.5	7218	1	US-08-232-463-14
7	34.8	6.9	7218	1	US-08-232-463-14
8	32.6	6.4	3471	2	US-08-715-568A-2
9	31.8	6.3	80246	4	US-09-078-294-4
10	31.8	6.3	80595	4	US-09-078-294-3
11	31.6	6.2	978	3	US-09-267-031-13
12	30	5.9	5962	6	5386025-5
13	30	5.9	5975	1	US-08-404-354B-1
14	30	5.9	5975	1	US-08-314-083B-1
15	30	5.9	5975	1	US-08-435-675B-1
16	30	5.9	5975	1	US-08-336-257A-3
17	30	5.9	5975	3	US-08-884-599-1
18	29.4	5.8	2338	2	US-08-425-069-1
19	29.4	5.8	2338	2	US-08-317-844B-1
20	29.4	5.8	3066	4	US-09-086-912-1
21	29	5.7	289	4	US-09-007-005-17
22	29	5.7	289	4	US-09-244-796-17
23	29	5.7	72928	3	US-09-009-913-1
24	28.6	5.6	87350	3	US-08-781-891-79
25	28.4	5.6	545	4	US-09-073-297-17
26	28.4	5.6	3489	2	US-08-728-323A-1
27	28.4	5.6	4181	1	US-07-670-611-1

Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 20, Appli
Sequence 68, Appli
Sequence 68, Appli
Sequence 68, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 6, Appli
Sequence 25, Appli
Sequence 11, Appli
Sequence 11, Appli
Sequence 11, Appli
Sequence 64, Appli

ALIGNMENTS

RESULT 1
US-08-964-725-4
; Sequence 4, Application US/08964725
; Patent No. 5939265
; GENERAL INFORMATION:

APPLICANT: COHEN, Maurice
APPLICANT: FRIEDMAN, Paula N.
APPLICANT: GORDON, Julian
APPLICANT: HODGES, Steven C.
APPLICANT: KLASS, Michael R.
APPLICANT: KRATOCHVIL, Jon D.
APPLICANT: ROBERTS-RAPP, Lisa
APPLICANT: RUSSELL, John C.
APPLICANT: STROUPE, Steven D.

TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
FOR DETECTING DISEASES OF THE LUNG
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA

ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA: US/08/964,725

APPLICATION NUMBER: 5939265

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Becker, Cheryl L.

REGISTRATION NUMBER: 35,441

REFERENCE/DOCKET NUMBER: 5997 US.P.1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 847/935-1729

TELEFAX: 847/938-2623

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 507 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

5939265

US-08-964-725-4

Query Match 100.0%; Score 507; DB 2; Length 507;
Best Local Similarity 100.0%; Pred. No. 2.6e-147;
Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagtggaaccactggctgggtgatttctgtagatttttctgatttttaaaactctg 60
Db 1 ggcagtggaaccactggctgggtgatttctgtagatttttctgatttttaaaactctg 60

QY 61 aaaaatccagataactgcatgaagctggttaactatcttctgctggtgaccatcag 120
Db 61 aaaaatccagataactgcatgaagctggttaactatcttctgctggtgaccatcag 120

QY 121 cctttagttactctgctactgctctctctcatcaaaaagtgcccttctctgttgcaa 180
Db 121 cctttagttactctgctactgctctctctcatcaaaaagtgcccttctctgttgcaa 180

QY 181 gtggaaccttaactctggagaacattcttccctttatggtatcattaaagcttctct 240
Db 181 gtggaaccttaactctggagaacattcttccctttatggtatcattaaagcttctct 240

QY 241 gaaaactctggcattctctgtgaagccttctgagccttctgagggcctgaagagtgataatga 300
Db 241 gaaaactctggcattctctgtgaagccttctgagccttctgagggcctgaagagtgataatga 300

QY 301 gctgggaccagagcttctgaaactgtgaagaaactgctggagcgcctatcacacttgg 360
Db 301 gctgggaccagagcttctgaaactgtgaagaaactgctggagcgcctatcacacttgg 360

QY 361 gtacatcaagataaagcggaggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 420
Db 361 gtacatcaagataaagcggaggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 420

QY 421 tgcgtgaacctgtctaccatattagatcaaatgcccctaaagttagtgaccctgaa 480
Db 421 tgcgtgaacctgtctaccatattagatcaaatgcccctaaagttagtgaccctgaa 480

QY 481 aagacaataaagcaatgaatacatt 507
Db 481 aagacaataaagcaatgaatacatt 507

RESULT 2

US-08-964-725-5
; Sequence 5, Application US/08964725
; Patent No. 5939265

GENERAL INFORMATION:

APPLICANT: COHEN, Maurice
APPLICANT: FRIEDMAN, Paula N.
APPLICANT: GORDON, Julian
APPLICANT: HODGES, Steven C.
APPLICANT: KLASS, Michael R.
APPLICANT: KRATOCHVIL, Jon D.
APPLICANT: ROBERTS-RAPP, Lisa
APPLICANT: RUSSELL, John C.
APPLICANT: STROUPE, Steven D.
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
FOR DETECTING DISEASES OF THE LUNG
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964,725
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 5997.US.P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 519 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-964-725-5

Query Match 98.5%; Score 499.4; DB 2; Length 519;
Best Local Similarity 99.8%; Pred. No. 5.8e-145;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 gtggaaccactggctgggtgatttctgtagatttttctgatttttaaaactctgaaaa 65
Db 19 GTGGAACCACTGGCTGGTGGATTTCGTAGATTTTCTGATTTTAAACTCTCTGAAAA 78

QY 66 tatccagataactgcatgaagctggttaactatcttctctgctggtgaccatcagcctt 125
Db 79 TATCCAGATAGTGTGATGAAGCTGGTAACATATCTTCTGCTGGTGACCATCAGCCTTT 138

QY 126 gtatttactctactgctctctctctctcaacaaagtgcctctctctgttgacaagtgg 185
Db 139 GTAGTACTCTGCTACTGCTTCTCTCATCAACAAAGTGCCCTTCTGTTGACAAGTTGG 198

QY 186 cactttactctggacaacattcttctctctctctctctctctctctctctctctctct 245
Db 199 CACCTTACTCTGGACAACATCTTCCCTTATGATCCATTAAGCTTCTCTGAAAA 258

QY 246 ctctgggcaattctgttgagcacttctgtgagggctgaggggctgaaggaagtgtgtaagtgg 305
Db 259 CTCTGGCATTCTGTGAGCACCTTGTGGAGGGCTAAGGAAGTGTGTAAATGAGCTGG 318

QY 306 gaccagagcttctgaagctgtgaagaaactctgagggcgctatcacacttgggtgac 365
Db 319 GACCAGAGCTTCTGAAGCTGTGAAGAACTGCTGGAGGCGCTATCACACTTGGTGTGAC 378

QY 366 atcaagataaagcggaggtggtgagggatggaagatgagctctctctctctctctctct 425
Db 379 ATCAAGATAAGAGCGGAGGTGGATGGGATGGAAGATGATGCTCTATCTCTCCCTGCT 438

QY 426 gaaacctgttctaccaattatagatcaaatgcccctaaagttagtgacctggaagga 485
Db 439 GAAACCTGTCTACCAATTATAGATCAAAATGCCCTTAAATGTAGTGACCCCTGAAAAAG 498

QY 486 caataaagcaatgaatacat 506
Db 499 CAATAAAGCAATGAATACCT 519

RESULT 3

US-08-964-725-2
; Sequence 2, Application US/08964725
; Patent No. 5939265
; GENERAL INFORMATION:
APPLICANT: COHEN, Maurice
APPLICANT: FRIEDMAN, Paula N.
APPLICANT: GORDON, Julian

```

1  APPLICANT: HODGES, Steven C.
2  APPLICANT: KLASS, Michael R.
3  APPLICANT: KRATOCHVIL, Jon D.
4  APPLICANT: ROBERTS-RAPP, Lisa
5  APPLICANT: RUSSELL, John C.
6  APPLICANT: STROUPE, Steven D.
7  TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
8  TITLE OF INVENTION: FOR DETECTING DISEASES OF THE LUNG
9  NUMBER OF SEQUENCES: 19
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Abbott Laboratories
12 CITY: Abbott Park
13 STATE: IL
14 COUNTRY: USA
15 ZIP: 60064-3500
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Diskette
18 COMPUTER: IBM Compatible
19 OPERATING SYSTEM: DOS
20 SOFTWARE: FastSeq for Windows Version 2.0
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/08/964,725
23 FILING DATE:
24 CLASSIFICATION: 435
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER:
27 FILING DATE:
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Becker, Cheryl L.
30 REGISTRATION NUMBER: 35,441
31 REFERENCE/DOCKET NUMBER: 5997.US.P1
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: 847/935-1729
34 - TELEFAX: 847/938-2623
35 TELEX:
36 INFORMATION FOR SEQ ID NO: 2:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 263 base pairs
39 TYPE: nucleic acid
40 STRANDEDNESS: single
41 TOPOLOGY: linear
42 US-08-964-725-2

```

```
Query Match          51.9%; Score 263; DB 2; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.1e-72;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	136	tgtctatgccttctctcatcaacaaagtcgcccttctcgttgacaagttggcaccctttacc	195
Db	1	TGCTACTGCTTCTCTCATCAACAAAGTGCCCTTCTCTGTGACAAGTTGGCACCTTTTACC	60
QY	196	tctggacaacattcttccctttatgatccattaaagcttcttctgaaacctctggcat	255
Db	61	CTTGGACACATTTCTTCCCTTTATGGATCCATTTAAAGCTTCTTCTGAAACCTCTGGGCAT	120
QY	256	tctcttgtagcaccttgtggaggggctaaggagtggtgtaaatgactgggaccagaggc	315
Db	121	TTCTGTGTGACACCTTGTGTGGGGGCTAAGGAAGTGTGTAATGAGCTGGGACCAGAGGC	180
QY	316	tcttgaagctgtgaagaactgctgagcgcgctatcacacttgggtgtgacatcaagataa	375
Db	181	TTCTTGAAGCTGTGAAGAACTGCTGGAGCGGCTTATCACACTTGGTGTGACATCAAGATAA	240
QY	376	agagcggaggtggatgggatgg	398
Db	241	AGAGCGGAGGTGGATGGGATGG	263

RESULT 4
US-08-964-725-1
; Sequence 1, Application US/08964725

```

: Patent No. 5939265
:
: GENERAL INFORMATION:
:
: APPLICANT: COHEN, Maurice
:
: APPLICANT: FRIEDMAN, Paula N.
:
: APPLICANT: GORDON, Julian
:
: APPLICANT: HODGES, Steven C.
:
: APPLICANT: KLASS, Michael R.
:
: APPLICANT: KRATOCHVIL, Jon D.
:
: APPLICANT: ROBERTS-RAPP, Lisa
:
: APPLICANT: RUSSELL, John C.
:
: APPLICANT: STROUPE, Steven D.
:
: TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
:
: TITLE OF INVENTION: FOR DETECTING DISEASES OF THE LONG
:
: NUMBER OF SEQUENCES: 19
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Abbott Laboratories
:
: STREET: 100 Abbott Park Road
:
: CITY: Abbott Park
:
: STATE: IL
:
: COUNTRY: USA
:
: ZIP: 60064-3500
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Diskette
:
: COMPUTER: IBM Compatible
:
: OPERATING SYSTEM: DOS
:
: SOFTWARE: FASTSEQ for Windows Version 2.0
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/964,725
:
: FILING DATE:
:
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER:
:
: FILING DATE:
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Becker, Cheryl L.
:
: REGISTRATION NUMBER: 35,441
:
: REFERENCE/DOCKET NUMBER: 5997.US.P1
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 847/935-1729
:
: TELEFAX: 847/938-2623
:
: TELEX:
:
: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 269 base pairs
:
: TYPE: nucleic acid
:
: STRANDEDNESS: single
:
: TOPOLOGY: linear
:
: US-08-964-725-1

```

	Query Match	51.9%	Score 263;	DB 2;	Length 269;
	Best Local Similarity	100.0%;	Pred. No. 4.2e-72;		
	Matches 263;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	6	gtggaaccactggcttggtagatttttgtagatattttctgatattttaaacctcctgaaaaa	65		
Dbb	7	GTGGAACCACTGGCTTGGTGATTTTGCCTAGATTTTTCTGATTTTAAACTCCTCAAAA	66		
QY	66	tatcccgagataacttgttcataagaagtggtaacctatatcttctcgtggtgacccaatcagccctt	125		
Dbb	67	TATCCCGAGATAAATGTCTATGAAGTGGTAACATATCTTCCTGCTGGTGACCATCAGCCCTTT	126		
QY	126	gtagttaactctgctactgcctctctcatcaacaagaagtcgccctctctctgttgacaagttgg	185		
Dbb	127	GTAGTTACTCTGCTACTGCCTTCTCATCAACAAGTSCCCCTTCCCTCTTGACAAGTTGG	186		
QY	186	caoctttaactctctgggaacaacattctccctttatgtgattccaataaaggcttctctgaaaaa	245		
Dbb	187	CACCTTTTACCCTCTGGACAACATCTTCCTTTATGGATTCCATTAAAGGCTTCTTCTGA	246		
QY	246	ctctggggcatttctgttgagcac	268		
Dbb	247	CTCTGGGCCATTTCTGTTGAGCAC	269		

	Query Match	44.4%;	Score 225;	DB 2;	Length 225;	
	Best Local Similarity	100.0%;	Pred. No. 2, 1e-60;			
	Matches :225;	Conservative	0;	Mismatches	0;	Indels
					0;	Gaps
283	aaggaagtgtgtaaatgactgggacagaggcttctgaagctgtgaagaaactgtgga	342				
1	AAGGAAAGTGTGTAATGAGCTGGCAGCAGAGGCTTCTGAAGCTGTGAAGAAACTGCTGGA	60				
343	ggcgctatcacacttggtgtgacatcaagataaagagcggaggtgtgattgggatggaaga	402				
61	GGCGCTATCACACTTGGTGTGACATCAAGATAAAGAGCGAGGTGGATGGGATGGAGA	120				
403	tgatgtccctatctccctccctggaacctgttctaccaattatagatcaaatgccttaa	462				
121	TGATGCTCCTATCCTCCCTGCCCTTGAACCTGTGTTCTACCAATTTATAGATCAAAATGCCCTAA	180				
463	aattgaatgaccctgtaaaaggagcaataaagaacatgaatacatt	507				

[illegible]

```

Qy 258 ctgttgagcacctt 271
: : : : :
Db 1288 YYYYYYYYYYYY 1301

RESULT 7
US-08-232-463-14/c
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: US/08/232.463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: ptzgpt-Fls
US-08-232-463-14

Query Match 6.9%; Score 34.8; DB 1; Length 7218;
Best Local Similarity 6.1%; Pred. NO. 0.6;
Matches 9; Conservative 91; Mismatches 48; Indels 0; Gaps 0;

Qy 274 ggaaggggctaaggagtgtgtaaatgagctgggaccagaggtcttctgaagctgtgaagaa 333
: : : : :
Db 1200 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1141

Qy 334 actgctgagcgctatcacactgggtgtgacatcaagataaagagcgaggtggaatggg 393
: : : : :
Db 1140 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1081

Qy 394 gatggaagatgatgtctctatctccct 421
: : : : :
Db 1080 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1053

RESULT 8

```

```

US-08-715-568A-2/c
: Sequence 2, Application US/08715568A
: Patent No. 5856463
: GENERAL INFORMATION:
: APPLICANT: Prydz, Hans Peter Blankenborg
: APPLICANT: Brede, Gaute
: TITLE OF INVENTION: PSKH-1 Ribozymes and Uses in Disease
: TITLE OF INVENTION: Treatment
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
: STREET: 600 South Avenue West
: CITY: Westfield
: STATE: NJ
: COUNTRY: USA
: ZIP: 07090-1497
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: US/08/715.568A
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Foley, Shawn P.
: REFERENCE/DOCKET NUMBER: FORSK 3.0-002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 908-654-5000
: TELEFAX: 908-654-7866
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3471 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
US-08-715-568A-2

Query Match 6.4%; Score 32.6; DB 2; Length 3471;
Best Local Similarity 51.7%; Pred. NO. 2;
Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 282 taagaaagtgtgtaaatgagctggagaccagagcttctgaagctgtgaagaaactgctgg 341
: : : : :
Db 1298 TATTGGAGCGTGTGGAGCGGCTGGAACCGGTGGACTGGGCAGATTGTGCTCTGGGCAGC 1239

Qy 342 aggcgtatcacacttgggtgtgacatcaagataaagagcgaggtggatgggagtggaag 401
: : : : :
Db 1238 GCGAGGAGGCACGTTTAAAGGAGGTTCTGGGATATGGACGCGTTCATGGATG 1179

Qy 402 atgatgtctctatctctccctgcc 424
: : : : :
Db 1178 AAGAGGAGCCATGCTCACCACC 1156

RESULT 9
US-09-078-294-4
: Sequence 4, Application US/09078294
: Patent No. 6265211
: GENERAL INFORMATION:
: APPLICANT: Choo, Kong-Hong Andy
: APPLICANT: Du Sart, Desiree
: APPLICANT: Cancilla, Michael R.
: TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
: FILE REFERENCE: Davies Col
: CURRENT APPLICATION NUMBER: US/09/078,294
: CURRENT FILING DATE: 1998-05-13
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 80246

```

```
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match
Best Local Similarity 6.3%; Score 31.8; DB 4; Length 80246;
Matches 66; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 117 teagcctttagtactctgctactgctctctctcatcaacaaagtgccttctgttg 176
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37301 tcaagtctcttttaattgaccacctctgctactctccacccttggacatctgctt 37360

QY 177 acaagttggacatttacctctggacaaattcttccctttatggtatccattaaagcttc 236
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37361 ctcaactggacactaccagtgctcaagattccctctcttttagtgatcttcagagcgc 37420

QY 237 ttc 239
|||
Db 37421 tac 37423

RESULT 10
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match
Best Local Similarity 6.3%; Score 31.8; DB 4; Length 80595;
Matches 66; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 117 teagcctttagtactctgctactgctctctctcatcaacaaagtgccttctgttg 176
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37563 tcaagtctcttttaattgaccacctctgctactctccacccttggacatctgctt 37622

QY 177 acaagttggacatttacctctggacaaattcttccctttatggtatccattaaagcttc 236
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37623 ctcaactggacactaccagtgctcaagattccctctcttttagtgatcttcagagcgc 37682

QY 237 ttc 239
|||
Db 37683 tac 37685

RESULT 11
US-09-267-031-13/c
; Sequence 13, Application US/09267031
; Patent No. 6137031
; GENERAL INFORMATION:
; APPLICANT: Zhang, Yuelin
; APPLICANT: Kinkema, Mark
; APPLICANT: Dong, Xinnian
; APPLICANT: Ronald, Pamela
; APPLICANT: Chern, Mawsheng
; APPLICANT: The Regents of the University of California
; APPLICANT: Duke University
; TITLE OF INVENTION: DNA Binding Proteins That Interact With NPRI

; FILE REFERENCE: 023070-0925000US
; CURRENT APPLICATION NUMBER: US/09/267,031
; CURRENT FILING DATE: 1999-03-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 978
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)-(978)
; OTHER INFORMATION: bZIP transcription factor gene TAG6
US-09-267-031-13

Query Match
Best Local Similarity 6.2%; Score 31.6; DB 3; Length 978;
Matches 49; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 95 actatctctgctggtgaccatcagcctttagttagtactgctactgcttccctc 154
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 207 ACTATCTTCAGCTGCTGAACATACGCTTCTTCCTCAATCTGCTTTCTTCTGCTGCTC 148

QY 155 acaaaagtgcctctctct 172
||| ||| ||| |||
Db 147 ACGATTTTGAGCAAGCCT 130

RESULT 12
5386025-5
; APPLICANT: JAY, SCOTT D.; ELLIS, STEVEN B.; HARPOLD, MICHAEL
; M.; CAMPBELL, KEVIN P.
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/482,384
; FILING DATE: 20-FEB-1990
; SEQ ID NO: 5
; LENGTH: 5962
5386025-5

Query Match
Best Local Similarity 5.9%; Score 30; DB 6; Length 5962;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 272 gtggaggggctaaaggaagtgtgtaaatgagctgggacagaggtcttctgaagctgtgaag 331
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3027 gtgcaggggctactactatgtgtacaaggacggggaccacacagatggagctgcgccc 3086

QY 332 aaactgctggagggcctacacttgggtgacatcaaat 373
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3087 cgcagtggtacacaaatgacttccactttgacacagtgtct 3128

RESULT 13
US-08-404-354B-1
; Sequence 1, Application US/08404354B
; Patent No. 5618720
; GENERAL INFORMATION:
; APPLICANT: Ellis, Steven Bradley
; APPLICANT: Williams, Mark E.
; APPLICANT: Harpold, Michael Miller
; APPLICANT: Schwartz, Arnold
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/314,083B
; FILING DATE: 28-SEPT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/914,231
; FILING DATE: 13-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 08-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-53191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 79...5700
; OTHER INFORMATION:
;
; US-08-314-083B-1
;
Query Match 5.9%; Score 30; DB 1; Length 5975;
Best Local Similarity 55.9%; Pred. No. 17;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0
Qy 272 gtggaggggctgaagaagtgtgtaaatgagctgggaccagaggcttctgaagctgtgaag 331
Db 2979 GTGCAGGGGGCTACTACTATGTGTACAAGGACGGGGACCCACGACGATGGAGCTGCGCCC 3038
Qy 332 aaactctgagggcgcgtatcacacttggtgtgacatcaagat 373
Db 3039 CCGCCAGTGGATACAAATGACTTCCTCCACTTTGACAACTGTCT 3080

RESULT 15
US-08-435-675B-1
; Sequence 1, Application US/08435675B
; Patent No. 5710250
; GENERAL INFORMATION:
; APPLICANT: Ellis, Steven Bradley
; APPLICANT: Williams, Mark E.
; APPLICANT: Harpold, Michael Miller
; APPLICANT: Schwartz, Arnold
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```

```
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,675B
FILING DATE: 05-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,083
FILING DATE: 28-SEP-1994
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
APPLICATION NUMBER: US 07/603,751
FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-53193
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
* NAME/KEY: Coding Sequence
* LOCATION: 79...5700
* OTHER INFORMATION:
US-08-435-675B-1
```

```
Query Match          5.9%; Score 30; DB 1; Length 5975;
Best Local Similarity 55.9%; Pred. No. 17;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 272 gtggaggggcctaaggaagtgtgtaaatgagctgggaccagagccttctgaagctgtgaag 331
    ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 2979 GTGCAGGGGCTACTATATGTGTACAGAGCGGGGACCCACGAGATGGAGCTGGCGCC 3038

Qy 332 aaactgctggaggcgctatcacacttggtgtgacatcaagat 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3039 CCGCCAGTGGATACACAATGACTTCCACTTTGACACGCTGCT 3080
```

Search completed: November 17, 2001, 14:47:27
Job time: 7125 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 12:37:21 ; Search time 2570.26 seconds
(without alignments)
1864.635 Million cell updates/sec

Title: US-09-700-770-1
Perfect score: 507

Sequence: 1 ggaagtggaaccactggct.....aataagcaatgaatacatt 507

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1:	gb_est1.*	44:	em_esthum10.*
2:	gb_est2.*	45:	em_esthum11.*
3:	gb_est3.*	46:	em_esthum12.*
4:	gb_est4.*	47:	em_esthum13.*
5:	gb_est5.*	48:	em_esthum14.*
6:	gb_est6.*	49:	em_esthum15.*
7:	gb_est7.*	50:	em_esthum16.*
8:	gb_est8.*	51:	em_esthum17.*
9:	gb_est9.*	52:	em_esthum18.*
10:	gb_est10.*	53:	em_esthum19.*
11:	gb_est11.*	54:	em_esthum20.*
12:	gb_est12.*	55:	em_esthum21.*
13:	gb_est13.*	56:	em_esthum22.*
14:	gb_est14.*	57:	em_esthum23.*
15:	gb_est15.*	58:	em_esthum24.*
16:	gb_est16.*	59:	em_esthum25.*
17:	gb_est17.*	60:	em_esthum26.*
18:	gb_est18.*	61:	em_esthum27.*
19:	gb_est19.*	62:	em_esthum28.*
20:	gb_est20.*	63:	em_estin1.*
21:	gb_est21.*	64:	em_estin2.*
22:	gb_est22.*	65:	em_estin3.*
23:	gb_est23.*	66:	em_estin4.*
24:	gb_est24.*	67:	em_estin5.*
25:	gb_est25.*	68:	em_estom1.*
26:	gb_est26.*	69:	em_estom2.*
27:	gb_est27.*	70:	em_estov1.*
28:	gb_est28.*	71:	em_estov2.*
29:	gb_est29.*	72:	em_estpl1.*
30:	gb_est30.*	73:	em_estpl2.*
31:	gb_est31.*	74:	em_estpl3.*
32:	gb_est32.*	75:	em_estpl4.*
33:	em_estba.*	76:	em_estpl5.*
34:	em_estfun.*	77:	em_estpl6.*
35:	em_esthum1.*	78:	em_estpl7.*
36:	em_esthum2.*	79:	em_estpl8.*
37:	em_esthum3.*	80:	em_estpl9.*
38:	em_esthum4.*	81:	em_estpl10.*
39:	em_esthum5.*	82:	em_estro1.*
40:	em_esthum6.*	83:	em_estro2.*
41:	em_esthum7.*	84:	em_estro3.*
42:	em_esthum8.*	85:	em_estro4.*
43:	em_esthum9.*	86:	em_estro5.*
		87:	em_estro6.*
		88:	em_estro7.*
		89:	em_estro8.*
		90:	em_estro9.*
		91:	em_estro10.*
		92:	em_estro11.*
		93:	em_estro12.*
		94:	em_estro13.*
		95:	em_estro14.*
		96:	em_estro15.*
		97:	em_estro16.*
		98:	em_estro17.*
		99:	em_estro18.*
		100:	em_estro19.*
		101:	em_estro20.*
		102:	gb_est25.*
		103:	gb_est26.*
		104:	gb_est27.*
		105:	gb_est28.*
		106:	gb_est29.*
		107:	gb_est30.*
		108:	gb_est31.*
		109:	gb_est32.*
		110:	gb_est41.*
		111:	gb_est42.*
		112:	gb_est43.*
		113:	gb_est44.*
		114:	gb_est45.*
		115:	gb_est46.*
		116:	gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB ID	Description
C 1	505	99.6	516	148	BF433853 7q58h12.x
C 2	503.4	99.3	549	155	BG537963 602563644
C 3	503	99.2	529	155	BG537963 602572977
C 4	494.4	97.5	509	137	BE550203 7b50b11.x
C 5	493	97.2	433	19	AI355302 qt57e07.x
C 6	490.4	96.7	825	155	BG540454 602568773
C 7	482	95.1	757	155	BG541018 602570177
C 8	479.2	94.5	825	154	BG484055 602504659
C 9	462.8	91.3	474	17	AI193988 qe81c04.x
C 10	458	90.1	461	144	BF111035 7n31d06.x
C 11	456.8	90.1	462	166	BE348591 ht72d05.x
C 12	452	89.2	817	155	BG547781 602575659
C 13	449.8	88.7	510	187	N92483 zb12g04.s1
C 14	446	88.0	464	14	AA992209 ot80f12.s
C 15	445	87.8	445	17	AI193016 qe69a09.x
C 16	444	87.6	450	19	AI355612 qt58b02.x
C 17	443	87.4	446	23	AI695085 we46f11.x
C 18	441	87.0	452	24	AI734244 zb12g04.y
C 19	440	86.8	453	23	AI668642 zb12g04.x
C 20	438	86.4	448	166	BE328852 hv97g12.x
C 21	434	85.6	435	18	AI263451 q107a12.x
C 22	433.2	85.4	438	18	AI299210 qn03d02.x
C 23	428	84.4	439	148	BF433537 7q54g01.x
C 24	424.4	83.7	702	154	BG507111 601861755
C 25	422	83.2	412	18	AI298558 qn02d04.x
C 26	416	82.1	466	14	AA988921 or81f01.s
C 27	399	78.7	405	19	AI354759 qt55d11.x
C 28	396	78.1	398	14	AA988559 or84f04.s
C 29	394	77.7	418	190	W40141 z074b09.r1
C 30	369.4	72.9	567	137	BE552278 hy05d09.x
C 31	360	71.0	362	164	BE222065 hu05d04.x
C 32	355.4	70.1	359	23	AI693419 wd90g12.x
C 33	352.4	69.5	356	166	BE326306 hv99c02.x
C 34	345.4	68.1	360	148	BF439805 nab58e04.
C 35	337.6	66.6	348	170	BF826945 CM4-HN002
C 36	321.4	63.4	733	154	BG482860 602502458
C 37	319	62.9	334	116	AW449467 U1-H-B13-
C 38	316.4	62.4	387	190	W17168 zb12g04.r1
C 39	299	59.0	311	18	AI265819 qx69a03.x
C 40	297.2	58.6	320	170	BF826946 CM4-HN002
C 41	295	58.2	310	156	D30977 HUML11966.H
C 42	286.4	56.5	290	156	D31546 HUML11712.Hu
C 43	282.6	55.7	333	156	D30961 HUML11877.H
C 44	281.4	55.5	299	156	D30979 HUML11982.H
C 45	279.6	55.1	291	156	D31534 HUML11612.Hu

ALIGNMENTS

RESULT 1
BF433853/c 516 bp mRNA EST 30-MAR-2001
LOCUS 7q58h12.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3702767 3',
DEFINITION mRNA sequence.
ACCESSION BF433853
VERSION BF433853.1 GI:11446183
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 516)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 481.
Location/Qualifiers
source 1. .516
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3702767"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"

FEATURES

note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 142 a 121 c 112 g 141 t
ORIGIN
Query Match 99.68; Score 505; DB 148; Length 516;
Best Local Similarity 100.0%; Pred. NO. 3.6e-141;
Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcagtggaaccactggcttggtgatttgcagattttctgatttttaaacctctg 60
DB 507 GGCAAGTGGACCACTGGCTGGTGGATTGCTAGATTTCTGATTTTAACTCTG 448
QY 61 aaaaatcccagataactgcatgaagctggtgaactatctctctggtgaccatcg 120
DB 447 AAAAATATCCAGATAAATGTCATGAAGCTGTAATCTCTGCTGTGACCATCAG 388
QY 121 cctttgtagttactctactgctctcctcctcatcaaaaagtcctctctgttgacaa 180
DB 387 CCTTTGTAGTTACTCTGCTACTGCTTCTCCTCATCAACAAAGTGCCCTTCTGTGACAA 328
QY 181 gttggcactttacctctggacaacattcttccctttatggtatcattaaagctcttct 240
DB 327 GTTGGACCTTTACCTCTGGACACATCTTCCCTTTATGGATCCATTAAGCTTCTCT 268
QY 241 gaaaactctggcatttcttggagcaaccttggaggggctgaaggaagtgttaaatga 300
DB 267 GAAACCTCTGGCATTTCTGTTGAGCACCTTGTGGAGGGGCTAAGGAAGTGTGTAATGA 208
QY 301 gctgggaccagagcttctgaagctgtgaagaaactgctggaggcgctatcacacttgg 360
DB 207 GCTGGGACCAAGAGGCTTCTTGAAGCTGTGAAGAAACTGCTGGAGGCGCTATCACACTTGG 148
QY 361 gtgacatcaagataaagagcgagggtggtggtggaagatgatgctctatctctccc 420
DB 147 GTGACATCAAGATAAAGACGAGGTGGATGGGATGGAAGATGATGCTCTCATCTCTCC 88
QY 421 tgcctgaacctgttctaccatattatagatcaaatgccctaaatagttagtaccctgaa 480
DB 87 TGCCTGAAACCTGTTCTACCAATATATAGATCAAAATGCCCTAAATGTAGTAGTACCTGCGTGA 28
QY 481 aaggacaaaataaagcaatgaatataca 505
DB 27 AAGGACAAAATAAGCAATGAATACA 3

```

RESULT 2
BG537963      549 bp      mRNA      EST      03-APR-2001
LOCUS      60256364F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4688315 5',
DEFINITION      mRNA sequence.
ACCESSION      BG537963
VERSION      BG537963.1 GI:13530195
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1502 row: h column: 12
High quality sequence stop: 515.
FEATURES
    source
        Location/Qualifiers
            1..549
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4688315"
                /clone_lib="NIH_MGC_77"
                /lab_host="DH10B (T1 phage-resistant)"
                /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                SfII (ggccattggcc); Site_2: SfII (ggccattggcc); 5' and
                3' adaptors were used in cloning as follows: 5' adaptor
                sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                5'-ATTCTAGAGCCGAGCGCCGACATG-dr(30)BN-3' (where B = A,
                C, or G and N = A, C, G, or T). Average insert size 1.9
                kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                by PCR. This library was enriched for full-length clones
                and was constructed by Clontech Laboratories (Palo Alto,
                CA). Note: this is a NIH_MGC Library."
BASE COUNT      162 a      121 c      119 g      147 t
ORIGIN
Query Match      99.3%; Score 503.4; DB 155; Length 549;
Best Local Similarity 99.8%; Pred. No. 1.1e-140;
Matches 504; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1      gcaagtgaaccactggtggtgatttgcataattttctgatttttaaacctctg 60
Db      14      GCGAAGTGGAACCACTGGCTTGGTGGATTTCCTAGATTTCCTGATTTTAAACTCTGT 73
QY      61      aaaaatccccagataactgtcatgaagctggtgaactatcttctgctggtgaaccatcag 120
Db      74      AAAAAATCCAGATAAATGTCATGAAGCTGTAATCTATCTTCCTGCTGGTGACCATCAG 133
QY      121      cctttgattactctgactgcttctctctatcaacaagaagtcgcccttctctgttgacaa 180
Db      134      CCTTTGATTACTCTGCTACTGCTTCTCTCATCAACAAAGTGCCCTTCCTGTGACAA 193
QY      181      gttggcacctttacctcttgacaaacattcttccctttatgaatccattaaagcttctct 240
Db      194      GTTGGACCTTTTACCTCTGGACAACATCTTCCCTTTATGATCCATTAAAGCTTCTTCT 253
QY      241      gaaactctggcattctgttgagcacctgtgtgagggggttaaggaagtgtgtaaatga 300
Db      254      GAAAACTCTGGCACTTCTGTGTGACACCTTCTGTGAGGGGCTAAGGAAGTGTGTAATGA 313

```

```

QY      301      gctgggaccagagcttctgaagctgtgaagaaactgctgagcgctatcacacttgg 360
Db      314      GCTGGGACCAGAGCTTCTGAAGCTGTGAAGAACTCTCTGGAGCGGTATCACACTTGGT 373
QY      361      gtgacatcaagataaaaagagcgaggtggtgagtgaggatggaagatgatgctcctccc 420
Db      374      GTGACATCAAGATAAAGAGCGGAGGTGGATGGGATGGAAGATGATGCTCTATCTCTCC 433
QY      421      tgctgaacacctgttctaccaattatagatcaaatgcctaaatgtagtgacccctgaa 480
Db      434      TGCTGAACACTGTCTTACCAATTATAGATCAAAATGCCCTAAATGTAGTGACCCCTGAA 493
QY      481      aaggacaaaataaagcaatgaataca 505
Db      494      AAGGACAAATAAAGCAATGAATAA 518

RESULT 3
BG545709      529 bp      mRNA      EST      04-APR-2001
LOCUS      60257297F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4701042 5',
DEFINITION      mRNA sequence.
ACCESSION      BG545709
VERSION      BG545709.1 GI:13544374
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1535 row: j column: 19
High quality sequence stop: 513.
FEATURES
    source
        Location/Qualifiers
            1..529
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4701042"
                /clone_lib="NIH_MGC_77"
                /lab_host="DH10B (T1 phage-resistant)"
                /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                SfII (ggccgctggcc); Site_2: SfII (ggccattggcc); 5' and
                3' adaptors were used in cloning as follows: 5' adaptor
                sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                5'-ATTCTAGAGCCGAGCGCCGACATG-dr(30)BN-3' (where B = A,
                C, or G and N = A, C, G, or T). Average insert size 1.9
                kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                by PCR. This library was enriched for full-length clones
                and was constructed by Clontech Laboratories (Palo Alto,
                CA). Note: this is a NIH_MGC Library."
BASE COUNT      149 a      114 c      118 g      148 t
ORIGIN
Query Match      99.2%; Score 503; DB 155; Length 529;
Best Local Similarity 100.0%; Pred. No. 1.4e-140;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      gcaagtgaaccactggtggtgatttgcataattttctgatttttaaacctctg 60
Db      13      GCGAAGTGGAACCACTGGCTTGGTGGATTTCCTAGATTTCCTGATTTTAAACTCTGT 72

```

Qy 61 aaaaataccagataactgtcatgaagctggtaactatcttctctgctggtgaccatcag 120
 Db 73 AAAAAATCCAGATAACTGTGATGAAGCTGGTAACATATCTTCTGCTGCTGACCATCAG 132
 Qy 121 cctttgtagttactgctactgctctccatcaacaagaagtgcccttctctgttgaca 180
 Db 133 CCTTGTAGTACTGCTACTGCTTCCCTCATCAACAAAAGTGCCCTTCTGTTGACAA 192
 Qy 181 gtgggaacctttaccttgacacaattcttcccttatggtatccattaaagctctctct 240
 Db 193 GTTGGCACCTTACCTCTGACACATCTTCCCTTATGATCCATTAAGCTTCTTCT 252
 Qy 241 gaaaactctggcattctctgttgacacctgtgaggggctgaaggaagtgtgtaataga 300
 Db 253 GAAAACTCTGGCATTCTCTGTAGCACCTTGTGGAGGGGCTAAGGAAGTGTGTAATGA 312
 Qy 301 gctgggacagagcttctgaagctgtgaagaaactctggaggcgctatcacacttggt 360
 Db 313 GCTGGGACCAAGAGCTTCTGAAGCTGTGAAGAACTCTCTGGAGCGCTATCACACTTGGT 372
 Qy 361 gtgacatacaagataaagacgaggtggtgaggtggaagatgagtctctctatccctccc 420
 Db 373 GTGACATCAGATAAAGAGCGGAGTGGATGGGATGGAAGATGATGCTCTATCTCTCC 432
 Qy 421 tgcctgaacctgtctaccatattatagatcaaatgccctaaatgtagtgtgacctgtaa 480
 Db 433 TGCTGAAACCTGTTCTACCAATTATAGATCAAAATGCCCTAAATGTAGTGACCCGTGAA 492
 Qy 481 aaggacaaaataaagcaatgaata 503
 Db 493 AAGGACAAATAAAGCAATGAATA 515

RESULT 4

BE550203/c 509 bp mRNA EST 10-AUG-2000
 LOCUS 7b50b11.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3231645 3',
 DEFINITION mRNA sequence.
 ACCESSION BE550203
 VERSION BE550203.1 GI:9791895
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 509)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
 Seq primer: -40UP from Gibco
 High quality sequence stop: 483.

FEATURES

source
 1..509
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3231645"
 /clone_lib="NCI_CGAP_Lu24"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; Plasmid DNA from the normalized
 library NCI_CGAP_Lu5 was prepared, and ss circles were

made in vitro. Following HAP purification, this DNA was
 used as tracer in a subtractive hybridization reaction.
 The driver was PCR-amplified cDNAs from a pool of 5,000
 clones made from the same library (cloneIDs
 1414920-1417991 and 1520904-1522439). Subtraction by Bento
 Soares and M. Fatima Bonaldo.
 BASE COUNT 145 a 115 c 113 g 136 t
 ORIGIN

Query Match 97.5%; Score 494.4; DB 137; Length 509;
 Best Local Similarity 99.8%; Pred. No. 5.6e-138;
 Matches 495; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 12 ccaactgcttggtgatttttctgtagatttttctgatttttaaaactcctgaaaaataatccc 71
 Db 500 CCACTGGCTTGGTGATTTTGTAGATTTTCTGATTTTAAACTCTCTGAAAAATATATCCC 441
 Qy 72 agataaactgtcatgaagctggttaactatcttctgctggtgacctcagcctttttagtt 131
 Db 440 AGATAACTGTGCATGAAGCTGGTAACATATCTTCTGCTGGTGACCATCAGCCTTTGTAGTT 381
 Qy 132 actctgactgctctcctcatcaacaagaagtgcccttctctgttgacaagtggcaacctt 191
 Db 380 ACTCTGTACTGCTTCTCTCATCAACAAGTGCCCTTCTCTGTTGACAAGTTGGACCTT 321
 Qy 192 tactctggacaacattcttccctttatggtatccattaaagctctctctgaaaaactctgg 251
 Db 320 TACCTCTGGACAACATTTCTCCCTTTATGGATCCATTAAAGCTTCTTCTGAAAACTCTGG 261
 Qy 252 gcattctgttgagcaccttggagggtgaggaagtgtgaaatgagctgggaccag 311
 Db 260 GCATTTCTGTTGAGCACCTTGTGGAGCGGCTAAGGAAGTGTGTAATGAGCTGGGACCAG 201
 Qy 312 aggtcttgaaactgtgaagaaactgctggaggcgctatcacacttggtgacatacaag 371
 Db 200 AGGCTTCTGAAGCTGTGAAGAAACTGCTGGAGCGCTATCACACTTGGTGTGACATCAAG 141
 Qy 372 ataaagagcgagtgatggggatggaagatgatcctctatcctcctcctcctggaacc 431
 Db 140 ATAAAGAGCGGAGGTGATGGGATGGAAGATGATGCTCTCTCTCTCCCTGCTGAAACC 81
 Qy 432 tgttctaccattatagatcaaatgccctaaatgtagtgcacctgtaaaaggacaaata 491
 Db 80 TGTCTTACCAATTATAGATCAAAATGCCCTAAATGTAGTGACCCCTGAAAAAGGACAAATA 21
 Qy 492 aagcaatgaatacatt 507
 Db 20 AAGCAATGAATACATT 5

RESULT 5

AI355302/c 493 bp mRNA EST 04-JAN-1999
 LOCUS qt57e07.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
 DEFINITION IMAGE:1952100 3', mRNA sequence.
 ACCESSION AI355302
 VERSION AI355302.1 GI:4095455
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 493)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco

[illegible]

Qy	181	gttggaacatttaacctgtgagaacaattcttccctttataggatccattaaagcttctctt	2400
Db	193	gttggcaccctttacctctggacaacattcttccctttatggatccattaaagcttctctt	2520
Qy	241	gaaactctggcattctctgttgagcacctctgtgagggctgaaggaagtggtaaatga	3000
Db	253	gaaactctggcattctctgttgagcacctctgtgagggctgaaggaagtggtaaatga	3120
Qy	301	gctgggaccagaggctcttgaagctgtgaagaacctgctggaggcgctatcacacttgg	3600
Db	313	gctgggaccagaggctcttgaagctgtgaagaacctgctggaggcgctatcacacttgg	3720
Qy	361	gtgacatcaagataaagacgaggtgagatgggagatgaagatgagatcctcatctccc	4200
Db	373	gtgacatcaagataaagacgaggtgagatgggagatgaagatgagatcctcatctccc	4320
Qy	421	tgctgaaacctgtctaccataattatagatcaaatgccctaaaatgtag-tgacccgtga	4790
Db	433	tgctgaaacctgtgtctaccataattatagatcaaatgccctaaaatgtagtgacccgtga	4920
Qy	480	a-aaggacaaataaagcaatgaataac 504	
Db	493	acaggacaaataaagcaatgaataac 518	
RESULT	8		

```

RESULT      8
BG484055   825 bp      mRNA      21-MAR-2001
LOCUS      602504659f1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4617999 5
DEFINITION mRNA sequence.
ACCESSION  BG484055
VERSION    BG484055.1 GI:13416334
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 825)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LUCM1375 row: f column: 16
            High quality sequence stop: 491.
            Location/Qualifiers
                1..825
FEATURES
source

```

```

/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site:1:
SfiI (ggcgctctggcc); Site:2: SfiI (ggccattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor sequence:
5'-ATTCTAGGCGGAGGGCGGCACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
221 a 202 c 192 g 210 t

```

```

Query Match      94.5%; Score 479.2; DB 154; Length 825;
Best Local Similarity 98.0%; Pred. No. 2.5e-133;
Matches 496; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

QY 1 ggcagtggaaccactgcttgatggtttgttagatattttctgatttttaaacctcgtg 60
Db 13 GGCAGTGGAAACCACTGGCTGTGTGGATTTGTAGATTTTCTGATTTTAAACFCCG 72
QY 61 aaaaatccccagataactgctatgaagctggtaacctgttaacctatctctctgctgtaaccatcg 120
Db 73 AAAAATATCCAGATAACTGTCATGAAGTGGTAACATATCTCTCTGCTGGTGACCATGAG 132
QY 121 cctttgtagttactctgactactcctctctctctctctctctctctctctctctctctctctct 180
Db 133 CCTTTGTAGTTACTCTGCTACTGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 192
QY 181 gttgcaactttacactctctggaacacattctccctttatggtaccattaaagcttctctct 240
Db 193 GTTGGCACCTTTACCTCTCGAACAACATCTTCCCTTTATGGATCCATTAAAGCTTCTTCT 252
QY 241 gaaactctgggcattctctgttgagcacctgttgaggggcttaaggaagtgtgtaaatga 300
Db 253 GAAACTCTGGGCATTCTCTGTGACACCTTGTGGAGGGGCTTAAGGAAGTGTGTAACATGA 312
QY 301 gctgggaccagaggtctctgaaagctgtgaagaacctgctggaggcgctatcacacttgggt 360
Db 313 GCTGGGACCAGAGGCTCTGAGCTGTGAAGAACTGCTGGAGGCGCTATCACACTTGGT 372
QY -361 gtgacatcaagaataagcgagggtgagtgaggatgg--aagatgatgctccctatctc 418
Db -373 GTGACATCAAGATCAGACGAGCGAGCTGGCTGGGATGACAGATGATGCTCTATCTCTC 432
QY 419 cctgctgaaactcttctacattatagatacaaaatgcccataaatgtagtgaccctg 478
Db 433 CCTGCTGAAACCTGTCTACCAATTATAGATCAAAATGCGCTTAAATGTATGACCGGTG 492
QY 479 aaaaagacaataaagaatgaatcac 504
Db 493 ACAGGCAAAATAAGCAATGAATAC 518

RESULT 9
AI193988/c
LOCUS
DEFINITION
IMAGE:1745382 3', mRNA sequence.
ACCESSION
AI193988
VERSION
EST.
KEYWORDS
human.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. 474
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1745382"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
FEATURES
source

```

strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NbHL19W.

BASE COUNT 130 a 103 c 106 g 135 t
ORIGIN

Query Match 91.3%; Score 462.8; DB 17; Length 474;
Best Local Similarity 99.6%; Pred. No. 1.8e-128;
Matches 464; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 40 ttcttgatttttaaacctcctgaaaaatatccagataaactgtcatgaaagctgtaactat 99
Db 474 TTTCTGATTTTAAATCCTCTGAAAAATATCCAGATAAAGTGTATGAGTGTGTAACATAT 415
QY 100 cttcctgctgggagccatcagcctttgttagttactctgctactgctctctctctcaacaa 159
Db 414 CTTCTCTGCTGGTGACCATCAGCCTTTGTAGTTACTCTGCTACTGCTCTCTCTCATCAACA 355
QY 160 agtgcctctctctgttgacaagtgtgcaacctttacctctggacaacattctctctttat 219
Db 354 AGTGCCCTCTCTGTTGTGACAAAGTTGGCACCITTAACCTCTGGACACATCTTCCCTTTAT 295
QY 220 ggatccattaaagctctcttgaaaaactctgggaactctgttgcacaccttggagggtgga 279
Db 294 GGATCCATTAAGCTTCTCTGAAAACTCTGGGCAATTTCTGTTGAGCACCTTGTGGAGGG 235
QY 280 gtaaggagtggtgtaaatgagctgggaccagaggtctctgaaagctgtgaaagaaactgct 339
Db 234 GCTAAGGAAGTGTGTAAATGAGCTGGGACCAAGAGCTTCTGAACTGTGAGAAACCTGTCT 175
QY 340 ggaggcgtatcacacttggtgtgacatcaagataaagcgagggtggtggtggtgga 399
Db 174 GGAGCGGTATCACACTTGTGTGTGACATCAAGATAAAGAGCGAGGTGGATGGGATGGA 115
QY 400 agatgctctctctctctctctgctgaaacctgttaccacattatagataaagctccc 459
Db 114 AGATGATGCTCTCTATCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 55
QY 460 taaaatgtagtaccctgaaagggacaataaagaatgaatcac 505
Db 54 TAAATGTAGTGACCCGCGTGAAGAGGACAAATAAGCAATGAATCCA 9

```

RESULT 10
BF111035/c
LOCUS
DEFINITION
IMAGE:1745382 3', mRNA sequence.
ACCESSION
BF111035
VERSION
EST.
KEYWORDS
human.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov

Seq primer: -40UP from Gbco.

FEATURES

source
1. .461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3566243"
/clone_lib="NCI-CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDs
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 126 a 102 c 103 g 129 t 1 others
ORIGIN

Query Match 90.3%; Score 458; DB 144; Length 461;
Best Local Similarity 99.8%; Pred. No. 4.9e-127;
Matches 458; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 47 tttttaactctgaaaaataccagataactgtcatgaagctggaactatctctctg 106
|||||
Db 461 TTTTAAACTCTGANAATATCCAGATAACTGTCATGAAGCTGGTAACATCTTCTG 402
Qy - 107 ctggtgaccatcagccttctgtagtactctgctactgctctctcctcacaacaaagtgc 166
|||||
Db 401 CTGGTGACCATCAGCCTTTGTAGTTACTCTGCTACTGCTCTTCCCTCATCAACAAAGTGCCC 342
Qy 167 ctctctgtgacaagtggacaccttacctctggacacattctctcttatgatcca 226
|||||
Db 341 CTTCCTGTGTGACAGTGGACACCTTACCTCTGGACAACTTCTCCCTTTATGGATCCA 282
Qy 227 ttaagctctcttgaaaactctggcattctctgtgagcaccttctgaggggctaaagg 286
|||||
Db 281 TTAAGCTTCTCTGAAACTCTGGCATTTCTGTGAGCACCTTGTGAGGGGCTAAGG 222
Qy 287 aagtgttaaatgagctgggaccagagctctctgaagctgtgaagaaactgctggagggc 346
|||||
Db 221 AAGTGTGTAATGAGCTGGGACCAAGAGCTTCTGAAGCTGTGAAGAACTGCTGGAGGCG 162
Qy 347 ctatcacacttggtgacatcaagataaagcagagtgatggagtgatggaagatgat 406
|||||
Db 161 CTATCACACTTGGGTGACATCAAGATAAAGCGGAGGTGGATGGGATGGAAGATGAT 102
Qy 407 gctcctatctccctgctgaaactctgtacctaatatgatgcaaatgcccctaaaatg 466
|||||
Db 101 GCTCCTATCCTCCCTGCTGAACTGTTCTTACCAATTTATAGATCAATGCCCTAAATATG 42
Qy 467 tagtgaccctgtgaaaaggacaaaataaagcaatgaatata 505
|||||
Db 41 TAGTGACCCGTGAAAGGACAAATAAAGCAATGAATACA 3

RESULT 11
BE348591/c
LOCUS
DEFINITION BE348591 462 bp mRNA EST 18-JUL-2000
hc72d05.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3152265 3',
mRNA sequence.
ACCESSION BE348591
VERSION BE348591.1 GI:9260444
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 462)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL

Tumor Gene Index

COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

FEATURES

Seq primer: -40UP from Gbco.

Location/Qualifiers

source

1. .462

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3152265"

/clone_lib="NCI-CGAP_Lu24"

/tissue_type="carcinoid"

/lab_host="DH10B"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP_Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.

The driver was PCR-amplified cDNAs from a pool of 5,000

clones made from the same library (cloneIDs

1414920-1417991 and 1520904-1522439). Subtraction by Bento

Soares and M. Fatima Bonaldo."

BASE COUNT

127 a 102 c 103 g 130 t

ORIGIN

Query Match 90.1%; Score 456.8; DB 166; Length 462;
Best Local Similarity 99.6%; Pred. No. 1.1e-126;
Matches 458; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 46 atttttaactctgaaaaataccagataactgtcatgaagctggaactatctct 105
|||||
Db 462 ATTTTAAACTCTCTGAAAAATATCCAGATAACTGTCATGAAGCTGGTAACATCTTCT 403
Qy 106 gctggtgaccatcagccttctgtagtactctgctactgctctctctcctcacaagaagtgc 165
|||||
Db 402 GCTGTGACCATCAGCCTTTGTAGTTACTTGTCTACTGCTTCCCTCATCAACAAAGTGCC 343
Qy 166 ccttctctgtgacaagtggcacctttacctctctggacaaacattctcctttatggatcc 225
|||||
Db 342 CCTTCTCTTGACAAAGTTGGCACCTTTTACCTCTGGACAAACATTTCTCCTTTATGATGCC 283
Qy 286 attaaagctctctctgaaaaactctgggcatctctgttggacaccttctggaggggctaaag 285
|||||
Db 282 ATTAAGACTTCTCTGAAAACTCTGGGCATTTCTGTTGAGCACCTTGTGGAGGGGCTAAG 223
Qy 286 gaagtgtgtaaatgagctgggaccagagggcttctgaaagctgtgaagaactgctggagggc 345
|||||
Db 222 TAAGTGTGTAATGAGCTGGGACCAAGAGGCTTCTGAAGCTGTGAAGAAACTGCTGGAGGC 163
Qy 346 gctatcacacttggtgtgacatcaagataaagagcggagggtggatggggatggaagatga 405
|||||
Db 162 GCTATCACACTTGGTGTGACATCAAGATAAAGAGCGGAGGTGGATGGGATGGAAGATGA 103
Qy 406 tgcctctatcctcctgctgctgaaacctgtctacaaattatagatcaaatgctccctaaaat 465
|||||
Db 102 TGCTCTATCTCTCCCTGCCTGCTTACCACTTTATAGATCAATGCTCCCTTAAAT 43
Qy 466 gtagtgaacccgtgaaaaggacaaaataaagcaatgaatata 505

[illegible]

```
Matches 482; Conservative 0; Mismatches 17; Indels 2; Gaps 2;
Qy 6 gtgaacacactggttggtgatttctgtagattttcttctgtagattttttaaaccctcgaataaa 65
Db 503 GGGGAACCACTGGCTGGCTGGATTTGCTAGTATTTCTGATTTTAACTCTCTGAAAAA 444
Qy 66 tatccagataactgctcat-gaaactggaactatctctcctgctggtgacacactcagcctt 124
Db 443 TATCCCAAGTAACGTCTCATGGAAGCTGGTAACATATCTTCTCTGCTGGTGAACCATCAGCCTT 384
Qy 125 tgaagtactctgctactgctctcctcctcctcctcctcctcctcctcctcctcctcctcct 184
Db 383 TGTAGTACTCTGCTACTGCTCTCTCTCATCAACAAGTGGCCCTCTCTCTGACAAAGTTG 324
Qy 185 gacactttacactcctggaacacattctctctctctctctctctctctctctctctctctct 244
Db 323 GCACCTTTACCTCTGGACACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGAAA 264
Qy 245 actctgggacttctgttgtagcaccttctgtaggggctcctcctcctcctcctcctcctcct 304
Db 263 ACTCTGGGCATTTCT-TTGAGCACCTTCTGGAGGGGCTAAGGAAGTGTGTAATGAGCTG 205
Qy 305 ggaccagagcttctggaactcctcctcctcctcctcctcctcctcctcctcctcctcctcct 364
Db 204 GGACCAGAGGCTCTGAAGCTGTGAAGAACTGCTGGAGGGGCTATCAACATTTGGTGTGA 145
Qy 365 catcaagataaagagcgagtgatggagtgatggaagatgatcctcctcctcctcctcctcct 424
Db 144 CATCAAGATAAAGAGCGAGGTGGATGGGATGGAGATGATGCTCTCTCTCTCTCTCTCTCT 85
Qy 425 tgaacctgttctaccatattatagatcaaatgcctcctcctcctcctcctcctcctcctcct 484
Db 84 TGAACCTGTTCTACCAATTTATAGATCAATGCTTAAATGTAGTACCCTGTGAAAGG 25
Qy 485 acaataaagaacatgaataca 505
Db 24 ACAAATAAAGCAATGAATACA 4

RESULT 14
AA992209/c 464 bp mRNA EST 27-AUG-1998
LOCUS ot80f12.1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
DEFINITION IMAGE:1623119 3', mRNA sequence.
ACCESSION AA992209
VERSION AA992209.1 GI:3178323
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 464)
AUTHORS NCI-CSAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 312 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 379.
FEATURES
Location/Qualifiers
1..464
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1623119"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/seq_stage="8-9 weeks"
/lab_host="DH10B"
/notice="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
```

was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - Oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGCGCGCTAAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 126 a 102 c 103 g 133 t
ORIGIN

Query Match 88.0%; Score 446; DB 14; Length 464;
Best Local Similarity 99.8%; Pred. No. 2e-123;
Matches 457; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy 48 ttttaaacctcgtgaaataatccagataaactgctcgaagctgtaactatctctcctgc 107
Db 461 TTTTAAACTCTGAAAAATATCCAGATAACTGTCATGAAGCTGGTAACATATCTTCTCTGC 402
Qy 108 tggtagacatcagccttctgtagtactcgtactcgtcctcctcctcctcctcctcctcct 167
Db 401 T-GTGACCACAGCTTTGTAGTTACTCTGCTACTGCTCTCTCTCTCTCTCTCTCTCTCT 343
Qy 168 ttctgttgacaagtggcacctttacctctggacaacattctcctcttcttggatccat 227
Db 342 TTCTCTGTTGACAAGTTGGCACCTTTACCTCTGGACAACATCTTCCCTTTATGGATCCAT 283
Qy 228 taaagcttctcctgaaactcctgggcatttctgttgagcaccttctgtaggggctaaaga 287
Db 282 TAAAGCTTCTTCTGAAAACTCTGGGCATTTCTGTTGAGCACCTTTGTGAGGGGCTTAAGA 223
Qy 288 agtggtataatgagctgggacagagcttctgaagctgtgaaagaacctgctggagcgcc 347
Db 222 AGTGTGTAATGAGCTGGGACACAGAGGCTTCTGAAGCTGTGAAGAACTGCTGGAGGCGC 163
Qy 348 tatcacacttggtgacatcaagataaagagcgagtgatggatggggatggaagatgatg 407
Db 162 TATCACACTTGGTGTGACATCAAGATAAAGAGCGGAGGTGGATGGGATGGAAGATGATG 103
Qy 408 ctctctatcctcctgctgaaacctgttctaccatattatagatcaaatgcctcctaaatgt 467
Db 102 CTCCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 43
Qy 468 agtgacctgtgaaagggacaaataaagcaatgaataca 505
Db 42 AGTGACCCGTGANAAGGACAATAAAGCAATGAATACA 5
```

```
RESULT 15
AA193016/c 445 bp mRNA EST 29-OCT-1998
LOCUS qe59a09.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:1744216 3', mRNA sequence.
ACCESSION AA193016
VERSION AA193016.1 GI:3744225
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 445)
AUTHORS NCI-CSAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 303 Std Error: 0.00
Seq primer: -40UP from Glibco.
Location/Qualifiers
```

```
source 1. .445
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:174216"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TCTTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Patima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W."

BASE COUNT 120 a 102 c 99 g 124 t
ORIGIN

Query Match 87.8%; Score 445; DB 17; Length 445;
Best Local Similarity 100.0%; Pred. No. 4e-123;
Matches 445; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 tgaataatccccgaataactgtcatgaagctggtgaactatctctctgtgtgaccatc 118
Db 445 TGAATAATATCCAGATAACTGTATGAAGCTGGTAACTATCTCTCTGTGTGACCATC 386
QY 119 agcctttgtagtctactgtactgtctctctcatcaaaagtgccccctctctgtgac 178
Db 385 AGCCTTTGTAGTACTCTGTACTGTCTCTCTCTCATCAAAAGTGCCCCCTCTCTGTGAC 326
QY 179 aagttggcacctttacctctgtgacaacattctccctttatggatccattaaagcttctt 238
Db 325 AAGTTGGCACCTTTTACCTCTGGACAACATCTTCCCTTTATGGATCCATTAAAGCTTCTT 266
QY 239 ctgaataactctggcattctgttgacacctgtgtgaggggctaaagagtggtgtaaat 298
Db 265 CTGAATAACTCTGGCATTCTGTGTGACACCTTGTGGAGGGGCTAAGGAAGTGTGTAAT 206
QY 299 gagctgggaccagggcttctgaagctgtgaagaaactgctggaggcgctatcacacttg 358
Db 205 GAGCTGGGACCAGAGGCTTCTGAAGCTGTGAAGAACTGCTGGAGGGCTATCACACTTG 146
QY 359 gtgtgacatcaagataaagcggaggtggatggggatggaagatgatgctctctctc 418
Db 145 GTGTGACATCAAGATAAAGAGCGGAGGTGGATGGGATGGAAGATGATGCTCTCTATCCTC 86
QY 419 cctgcctgaacctgttctaccattatagatcaaatgccctaaatgtagtgacccgtg 478
Db 85 CCTGCCTGAAACCTGTCTTACCAATTATAGATCAATGCCCTTAAATGTAGTGACCCGTG 26
QY 479 aaagacaataaagcaatgaata 503
Db 25 AAGAGGACAATAAAGCAATGAATA 1
```

Search completed: November 17, 2001, 13:33:57
Job time: 3396 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 12:51:12 ; Search time 222.76 Seconds
(without alignments)
1429.099 Million cell updates/sec

Title: US-09-700-770-1
Perfect score: 507
Sequence: 1 ggcaagtggaaaccactggct.....aataaagcaatgaatacatatc 507

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	507	100.0	507	AAV38069	Human lung tissue
2	507	100.0	507	AAV38069	LU103 specific pol
3	507	100.0	507	AAV38069	Human lung specific
4	506	99.8	531	AAV38069	Human signal pepti
5	499.4	98.5	519	AAV38069	Human lung tissue
6	499.4	98.5	519	AAV38069	LU103 specific pol
7	325.4	64.2	357	AAV86241	EST clone AA246
8	263	51.9	263	AAV38067	Human lung tissue
9	263	51.9	263	AAV38067	LU103 specific pol
10	263	51.9	269	AAV38066	Human lung tissue
11	263	51.9	269	AAV38066	LU103 specific pol

12	225	44.4	225	19	AAV38068	Human lung tissue
13	225	44.4	225	20	AAV38068	LU103 specific pol
14	58	11.4	60	16	AAV22859	Human gene signatu
15	56.4	11.1	225	19	AAV54618	LU105 specific pol
16	56.4	11.1	519	19	AAV54621	LU105 polypeptide
17	56.4	11.1	543	21	AAV298173	Human signal pepti
18	56.4	11.1	543	21	AAV29723	Human lung specific
19	56.4	11.1	562	19	AAV54620	LU105 specific con
20	56.4	11.1	570	21	AAV65103	Membrane-bound pro
21	56.4	11.1	570	22	AAV44249	Human PRO1245 (UNQ
22	45	8.9	936	22	AAV58252	Oligonucleotide D1
23	45	8.9	936	22	AAV58254	Oligonucleotide D1
24	45	8.9	936	22	AAV58257	Oligonucleotide D1
25	45	8.9	936	22	AAV58259	Oligonucleotide D2
26	45	8.9	936	22	AAV58262	Oligonucleotide D2
27	45	8.9	938	22	AAV58255	Oligonucleotide D1
28	42.2	8.3	936	22	AAV58252	Oligonucleotide D1
29	42.2	8.3	936	22	AAV58254	Oligonucleotide D1
30	42.2	8.3	936	22	AAV58257	Oligonucleotide D1
31	42.2	8.3	936	22	AAV58259	Oligonucleotide D1
32	42.2	8.3	936	22	AAV58262	Oligonucleotide D2
33	42.2	8.3	938	22	AAV58255	Oligonucleotide D1
34	34	6.7	1324	21	AAV45583	Arabidopsis thalia
35	34	6.7	1325	21	AAV33679	Arabidopsis thalia
36	34	6.7	1431	21	AAV46348	Arabidopsis thalia
37	34	6.7	1435	21	AAV40039	Arabidopsis thalia
38	33.4	6.6	1943	21	AAV47114	Arabidopsis thalia
39	33.4	6.6	5019	21	AAV45709	Arabidopsis thalia
40	32.6	6.4	3471	18	AAV71761	PSKH-1 cDNA. Homo
41	32.4	6.4	2815	20	AAV06817	Chlamydia pneumoni
42	32.4	6.4	2950	21	AAV30849	Chlamydia antigen
43	32.4	6.4	3000	20	AAV06828	Chlamydia pneumoni
44	32.4	6.4	8689	21	AAV97996	Human T gene cDNA.
45	32.4	6.4	8949	21	AAV07840	Human homologue of

ALIGNMENTS

RESULT 1
AAV38069
ID AAV38069 standard; cDNA; 507 BP.
XX AC AAV38069;
XX AC AAV38069;
DT 14-SEP-1998 (first entry)
XX DE Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:4.
XX DE Human lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.
XX KW Homo sapiens.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 83..364
XX FT /*tag= a
XX FT /product= "LU103"
XX PN WO9820143-A1.
XX PD 14-MAY-1998.
XX PF 05-NOV-1997; 97WO-US20680.
XX PR 05-NOV-1996; 96US-0744211.
XX PA (ABBO) ABBOTT LAB.
XX PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
XX PI Kratochvill JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX DR WPI; 1998-286957/25.
XX DR P-PSDB; AAV62068.

XX Lung tissue derived polynucleotide LU103 - useful to detect,
 PT diagnose, stage, monitor, prognosis, prevent, treat or determine
 XX pre-disposition to lung disease, e.g. lung cancer
 XX
 PS Claim 1; Page 67; 86pp; English.
 XX
 CC The present sequence represents a polynucleotide specific for lung
 CC tissue gene LU103. A method has been developed for detecting the
 CC presence of a target LU103 polynucleotide in a test sample, comprising:
 CC (a) contacting the sample with at least 1 LU103-specific polynucleotide,
 CC and (b) detecting the target LU103 polynucleotide in the test sample,
 CC where the LU103 polynucleotide has at least 50% identity to the 269,
 CC 263, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to
 CC AAV38070. The methods and products of the present invention may be used
 CC to detect, diagnose, stage, monitor, prognosis, prevent, treat or
 CC determine the predisposition diseases and conditions of the lung, e.g.
 XX lung cancer.
 XX
 SQ Sequence 507 BP; 137 A; 110 C; 116 G; 144 T; 0 other;

Query Match 100.0%; Score 507; DB 19; Length 507;
 Best Local Similarity 100.0%; Pred. No. 6.8e-156;
 Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcaagtggaccactggctgtggatttggtagatttttctgatttttaaaactctg 60
 Db - 1 ggcaagtggaccactggctgtggatttggtagatttttctgatttttaaaactctg 60
 QY 61 aaaaatccacagataactgcatgaagctggttaactattcttctggtgaccatcag 120
 Db - 61 aaaaatccacagataactgcatgaagctggttaactattcttctggtgaccatcag 120
 QY 121 cctttagttagtactgctactgctcttccatcaacaaagtgccttctctgttgacaa 180
 Db 121 cctttagttagtactgctactgctcttccatcaacaaagtgccttctctgttgacaa 180
 QY 181 gttggacccttaccctggacacatttccctttatgagatgacatttaaaagtcttct 240
 Db 181 gttggacccttaccctggacacatttccctttatgagatgacatttaaaagtcttct 240
 QY 241 gaaaactctggcattctgtgagacactgtggagggtcgaagaaagtgttaaatga 300
 Db 241 gaaaactctggcattctgtgagacactgtggagggtcgaagaaagtgttaaatga 300
 QY 301 gctgggaccagagcttctgaagctgtgaagaaactgctggaggcgtatcacacttgg 360
 Db 301 gctgggaccagagcttctgaagctgtgaagaaactgctggaggcgtatcacacttgg 360
 QY 361 gtgacatcaagataaagagcggggtggatgggagtggaagatgagctctatctccc 420
 Db 361 gtgacatcaagataaagagcggggtggatgggagtggaagatgagctctatctccc 420
 QY 421 tgctgaaacctgttaccataattatagatcaaatgcctctaaatgtagtaccocgtgaa 480
 Db 421 tgctgaaacctgttaccataattatagatcaaatgcctctaaatgtagtaccocgtgaa 480
 QY 481 aaggacaaataaagcaatgaatacatt 507
 Db 481 aaggacaaataaagcaatgaatacatt 507

RESULT 2
 AAX99424
 ID AAX99424 standard; DNA; 507 BP.
 XX
 AC AAX99424;
 XX
 DT 19-OCT-1999 (first entry)
 XX
 XX LU103 specific polynucleotide consensus sequence.

KW LU103; tumour; lung cancer; detection;
 KW Fluorescent in situ hybridisation; ss.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 85..364
 FT /*tag= a
 FT /product= "Sequence AAY28334 amino acid residue"
 PN US5939265-A.
 XX
 PD 17-AUG-1999.
 XX
 PF 05-NOV-1997; 97US-0964725.
 XX
 PR 05-NOV-1997; 97US-0964725.
 XX
 PR 05-NOV-1996; 96US-0744211.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
 PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
 XX
 DR WPI: 1999-468402/39.
 DR P-PSDB; AAY28334.
 XX
 PT Polynucleotides useful for detecting, diagnosing and monitoring
 PT diseases of the lung such as lung cancer
 PS
 PS Claim 1; Column 47-49; 36pp; English.
 XX
 CC The 507 base pair sequence of a LU103 specific polynucleotide was
 CC derived from a consensus of the isolated sequences from clones
 CC 1235095, 1235531 and 1379417.
 CC The polynucleotides and methods disclosed in the invention can be
 CC useful for detecting, diagnosing, staging, monitoring or predicting
 CC diseases and conditions of the lung, such as lung cancer.
 CC The polynucleotides may be used to produce probes for use in
 CC fluorescent in situ hybridization (FISH) technology to perform
 CC chromosomal analysis and identify cancer specific alterations
 CC such as deletions.
 XX
 SQ Sequence 507 BP; 137 A; 110 C; 116 G; 144 T; 0 other;

Query Match 100.0%; Score 507; DB 20; Length 507;
 Best Local Similarity 100.0%; Pred. No. 6.8e-156;
 Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcaagtggaccactggctgtggatttggtagatttttctgatttttaaaactctg 60
 Db 1 ggcaagtggaccactggctgtggatttggtagatttttctgatttttaaaactctg 60
 QY 61 aaaaatccacagataactgcatgaagctggttaactattcttctggtgaccatcag 120
 Db 61 aaaaatccacagataactgcatgaagctggttaactattcttctggtgaccatcag 120
 QY 121 cctttagttagtactgctactgcttccatcaacaaagtgccttctctgttgacaa 180
 Db 121 cctttagttagtactgctactgcttccatcaacaaagtgccttctctgttgacaa 180
 QY 181 gttggacccttaccctggacacatttccctttatgagatgacatttaaaagtcttct 240
 Db 181 gttggacccttaccctggacacatttccctttatgagatgacatttaaaagtcttct 240
 QY 241 gaaaactctggcattctgtgagacactgtggagggtcgaagaaagtgttaaatga 300
 Db 241 gaaaactctggcattctgtgagacactgtggagggtcgaagaaagtgttaaatga 300
 QY 301 gctgggaccagagcttctgaagctgtgaagaaactgctggaggcgtatcacacttgg 360
 Db 301 gctgggaccagagcttctgaagctgtgaagaaactgctggaggcgtatcacacttgg 360

Db	301	gtcggacacagagcctcttgaagcgtggaagaaactgctggaggcgctatcacacttgat	360
Qy	361	gtgacatcaagataaagagcggaggtggatggggatggaagatgatgctcctatctccc	420
Db	361	gtgacatcaagataaagagcggaggtggatggggatggaagatgatgctcctatctccc	420
Qy	421	tgcctgaaacctgttaccaaattatagatcaaatgccctaaatgtagtaccogtgaa	480
Db	421	tgcctgaaacctgttaccaaattatagatcaaatgccctaaatgtagtaccogtgaa	480
Qy	481	aaggacaaataaagcaatgaatacatt	507
Db	481	aaggacaaataaagcaatgaatacatt	507

RESULT 3

AAZ29721
ID AAZ29721 standard; DNA; 507 BP.

AAZ29721;

DT 27-MAR-2000 (first entry)

AA Human lung specific gene Lng101.

AA Lung Specific Gene; LSG; Lngl01; human; diagnostic marker;
KW prognosticate; lung cancer; diagnosis; ds.
KW

XX Homo sapiens.
OS

XX	Key	Location/Qualifiers
FH	CDS	83..364
FT		/*tag= a
FT		/product= "LSG Lngl

AX PN .WO9960160-A1.

XX - 25-NOV-1999.
PD

XX
PF 12-MAY-1999; 99WO-US10344.

XX 21-MAY-1998; 98US-0086212.

PA (DIAD-) DIADEXUS LLC.

XX
PI Yang F,
Macina RA, Sun Y;

AA
DR
WPI: 2000-116320/10.

DR P-PSDB; AAY44456.

PT A new method for diagnosing, monitoring and staging lung cancer

XX PS Claim 6; Page 33; 40pp; English.

The present sequence is a lung specific gene (LSG) (Lung101) from human clone ID 126758. The LSG has high level of tissue specificity for lungs and is overexpressed in cancerous tissues. The sequence serves as a diagnostic marker for detecting, monitoring, staging and prognosticating lung cancer. The diagnosis involves comparing levels of LSG in samples obtained from patient and normal control.

Sequence 507 BP; 137 A; 110 C; 116 G; 144 T; 0 other;
XX

Query Match 100.0%; Score 507; DB 21; Length 507;
Best Local Similarity 100.0%; Pred. No. 6.8e-156;
Matches 507; Conservative 0; Mismatches 0; Indels 0

Qy 1 ggcaagtgaaccactggccttggtggattttgctagatttttctgatttttaactcctg 60

Db 1 ggcaagtgaaccactggcttgggtggattttgctagatttttctgatttttaactcctg 60

Qy 61 aaaatatccagataaactgtcatgaagctggtaactatctcctgctggtgaccatcag 120

[illegible]

RESULT

AAZ98174
ID AAZ98174 standard; cDNA; 531 BP.

AA
AC AAZ98174;

11-MAY-2000 (first entry)

Human signal peptide containing protein HSPP-66 cDNA SEQ ID NO:200

xx Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
 kw inflammation; cardiovascular disease; anticancer; anti-inflammatory;
 kw antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
 kw antiasthmatic; gene therapy; cell proliferation; neurological disorder;
 kw reproductive disorder; developmental disorder; arteriosclerosis;
 kw arthritis; psoriasis; acquired immune deficiency syndrome; anaemia;
 kw asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
 kw Parkinson's disease; Huntington's diseases; ovulatory defect;
 kw muscular dystrophy; ss.

XX Homo sapiens. OS

XX PN WO200000610-A2.

XX PD 06-JAN-2000.

XX
PF 25-JUN-1999: 99WO-US14484.

XX
PR 26-JUN-1998: 98US-0090762.

PR 31-JUL-1998; 9805-0094983.
PR 01-OCT-1998; 9805-0102686

PR 11-DEC-1998; 98US-0112129.
XX

PA (INCY-) INCYTE PHARM INC.

PI Lal P, Tang YT, Gorgone

PI Bandman O;

DR WPI; 2000-

11/11/2011 11:11:11 AM

XX New human signal peptide-containing proteins useful in treatment,
 PT prevention and diagnosis of e.g. cancer, inflammation and
 PT cardiovascular disease
 XX
 PS Claim 9; Page 289; 327pp; English.
 XX
 CC AA298109 to AA298242 encode AY87224 to AY87357 which represent the
 CC human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have
 CC anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
 CC neuroprotective, cardiovascular and antiasthmatic activities, and can
 CC be used in gene therapy. HSPs can be used to treat or prevent disorders
 CC associated with decreased activity or function of HSP. Antagonists of
 CC HSP are used to treat or prevent disorders associated with increased
 CC activity or function of HSP. Such disorders include cell proliferation
 CC (including cancer), inflammation, cardiovascular, neurological,
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
 CC asthma, Crohn's disease, microbial or other infections, congestive or
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP
 CC nucleic acids can be used for the recombinant production of HSP, for
 CC detecting HSP in standard hybridisation and amplification assays (for
 CC triplex-forming or ribozyme therapeutics, in gene therapy, as antisense,
 CC or genetic variations, and for chromosomal mapping. HSP are also used to
 CC raise specific antibodies (Ab) and to screen for agonists and
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or
 CC monitor, HSP-related diseases (in usual immunoassays), as therapeutic
 CC antagonists, in competitive drug screens, and for purification of HSP
 CC from natural sources.
 XX
 SQ Sequence 531 BP; 149 A; 115 C; 119 G; 148 T; 0 other;

Query Match 99.8%; Score 506; DB 21; Length 531;
 Best Local Similarity 100.0%; Pred. No. 1.5e-155;
 Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcaagtgaaccactgcttggtgattgctagattttctgattttttaaactctga 61
 DB 1 gcaagtgaaccactgcttggtgattgctagattttctgattttttaaactctga 60
 QY 62 aaaaatccagataaactgctgaagctggttaactattcttctgctggtgaccatcag 121
 DB 61 aaaaatccagataaactgctgaagctggttaactattcttctgctggtgaccatcag 120
 QY 122 cttttagttactctgctactgcttctctcaatacaaaagtgcccttctctgttgacaag 181
 DB 121 cttttagttactctgctactgcttctctcaatacaaaagtgcccttctctgttgacaag 180
 QY 182 ttggacatttactcttggaacatttcttcttattgattccattaaagcttctcttg 241
 DB 181 ttggacatttactcttggaacatttcttcttattgattccattaaagcttctcttg 240
 QY 242 aaaaactctggcattctgttgaacaccttggaggggctgaagtggtgaaatgag 301
 DB 241 aaaaactctggcattctgttgaacaccttggaggggctgaagtggtgaaatgag 300
 QY 302 ctgggaccagaggtcttctgaagctggaagaactgctgagcgctacacacttggtg 361
 DB 301 ctgggaccagaggtcttctgaagctggaagaactgctgagcgctacacacttggtg 360
 QY 362 tgacatcaagataaagagcgagggtgagtgggatggaagatgatgctctatccct 421
 DB 361 tgacatcaagataaagagcgagggtgagtgggatggaagatgatgctctatccct 420
 QY 422 gctgaacctgtctaccatattagatcaaatgcccataaagttagtaccctgtgaa 481
 DB 421 gctgaacctgtctaccatattagatcaaatgcccataaagttagtaccctgtgaa 480
 QY 482 aggaacaataaagcaatgaatcatt 507
 DB 482 aggaacaataaagcaatgaatcatt

Db 481 aggaacaataaagcaatgaatcatt 506
 RESULT 5
 AAV38070
 ID AAV38070 standard; cDNA; 519 BP.
 XX
 AC AAV38070;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:5.
 XX
 KW Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 96..377
 FT /*tag= a
 FT /product= "LU103"
 XX
 PN WO9820143-A1.
 XX
 PD 14-MAY-1998.
 XX
 PF 05-NOV-1997; 97WO-US20680.
 XX
 PR 05-NOV-1996; 96US-0744211.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
 PI Kratochvill JD, Roberts-Rapp L, Russell JC, Stroupe SD;
 XX
 DR WPI: 1998-286957/25.
 DR P-PSDB; AAW62068.
 XX
 PT Lung tissue derived polynucleotide LU103 - useful to detect,
 PT diagnose, stage, monitor, prognosis, prevent, treat or determine
 PT pre-disposition to lung disease, e.g. lung cancer
 XX
 PS Claim 1; Page 68; 86pp; English.
 XX
 CC The present sequence represents a polynucleotide specific for lung
 CC tissue gene LU103. A method has been developed for detecting the
 CC presence of a target LU103 polynucleotide in a test sample, comprising:
 CC (a) contacting the sample with at least 1 LU103-specific polynucleotide,
 CC and (b) detecting the target LU103 polynucleotide in the test sample,
 CC where the LU103 polynucleotide has at least 50% identity to the 269,
 CC 263, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to
 CC AAV38070. The methods and products of the present invention may be used
 CC to detect, diagnose, stage, monitor, prognosis, prevent, treat or
 CC determine the predisposition diseases and conditions of the lung, e.g.
 CC lung cancer.
 XX
 SQ Sequence 519 BP; 140 A; 113 C; 117 G; 149 T; 0 other;

Query Match 98.5%; Score 499.4; DB 19; Length 519;
 Best Local Similarity 99.8%; Pred. No. 2.1e-153;
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 6 gtggaaccactggttggtgatttctgtagattttctgatttttaaacctcctgaaaaa 65
 DB 19 gtggaaccactggttggtgatttctgtagattttctgatttttaaacctcctgaaaaa 78
 QY 66 tatccagataaactgtcatgaagctggaactatcttctgctgggacacatcagcttt 125
 DB 79 tatccagataaactgtcatgaagctggaactatcttctgctgggacacatcagcttt 138
 QY 126 gtagtactctgactgcttctcctcatcaacaagaagtccttctctgttgacaagtgg 185
 DB 126 gtagtactctgactgcttctcctcatcaacaagaagtccttctctgttgacaagtgg

Db 139 gtagttactctgtactgtccttccctcatcaacaagtgcccttctctgttgacaagtgg 198
QY 186 cacccttaccctctggacaacattctccctttatggtatccattaaagctctctctgaaaa 245
Db 199 cacccttaccctctggacaacattctccctttatggtatccattaaagctctctctgaaaa 258
QY 246 cctctggcattctctgtgagcaccctctgtgaggggctaaaggaagtgtgtaaatgagctgg 305
Db 259 cctctggcattctctgtgagcaccctctgtgaggggctaaaggaagtgtgtaaatgagctgg 318
QY 306 gaccagagctctgaagctgtgaagaaactctggaagcgtatcacacttgggtgac 365
Db 319 gaccagagctctgaagctgtgaagaaactctggaagcgtatcacacttgggtgac 378
QY 366 atcaagataaagcggaggtgagtggaagatgatgctctatctctccctgcct 425
Db 379 atcaagataaagcggaggtgagtggaagatgatgctctatctctccctgcct 438
QY 426 gaaacctgttctaccattatagatcaaatgccctaaaaatgtagtaccctgtgaaaagga 485
Db 439 gaaacctgttctaccattatagatcaaatgccctaaaaatgtagtaccctgtgaaaagga 498
QY 486 caaataaagcaatgaatacat 506
Db 499 caaataaagcaatgaatacct 519
RESULT 6
AAX99425
ID AAX99425 standard; DNA; 519 BP.
XX
AC AAX99425;
XX
DT 19-OCT-1999 (first entry)
XX
DE LU103 specific polynucleotide.
XX
KW LU103; tumour; lung cancer; detection; FISH;
KW Fluorescent in situ hybridisation; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN US5939265-A.
XX
PD 17-AUG-1999.
XX
PF 05-NOV-1997; 97US-0964725.
XX
PR 05-NOV-1997; 97US-0964725.
PR 05-NOV-1996; 96US-0744211.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX
DR WPI; 1999-468402/39..
XX
PT Polynucleotides useful for detecting, diagnosing and monitoring
PT diseases of the lung such as lung cancer
XX
PS Claim 1; Column 49; 36pp; English.
XX
CC The 519 base pair sequence of a LU103 specific polynucleotide.
CC The polynucleotides and methods disclosed in the invention can be
CC useful for detecting, diagnosing, staging, monitoring or predicting
CC diseases and conditions of the lung, such as lung cancer.
CC The polynucleotides may be used to produce probes for use in
CC fluorescent in situ hybridization (FISH) technology to perform
CC chromosomal analysis and identify cancer specific alterations
CC such as deletions.
XX

SQ Sequence 519 BP; 140 A; 113 C; 117 G; 149 T; 0 other;
Query Match 98.5%; Score 499.4; DB 20; Length 519;
Best Local Similarity 99.8%; Pred. No. 2.1e-153;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6 gtggaacacactggtggtgattttgctagatttttctgatttttaaacctcctgaaaaa 65
Db 19 gtggaacacactggtggtgattttgctagatttttctgatttttaaacctcctgaaaaa 78
QY 66 tatccacagataaactgtoatgaagctggttaactatctctcctgctggtgaccatcagccttt 125
Db 79 tatccacagataaactgcatgaagctggttaactatctctcctgctggtgaccatcagccttt 138
QY 126 gtagtctactctgactgcttctcctcatcaacaagtgcccccttctctgttgacaagtgg 185
Db 139 gtgttactctgctactgcttctcctcatcaacaagtgcccccttctctgttgacaagtgg 198
QY 186 cactttacctctggacaacattctctcctttatggatccattaaagctctctctgaaaa 245
Db 199 cactttacctctggacaacattctctcctttatggatccattaaagctctctctgaaaa 258
QY 246 cctctggcattctctgtgagcaccctctgtgaggggctaaaggaagtgtgtaaatgagctgg 305
Db 259 cctctggcattctctgtgagcaccctctgtgaggggctaaaggaagtgtgtaaatgagctgg 318
QY 306 gaccagagctctctgaagctgtgaagaaactctgctgagcgctatcacacttgggtgac 365
Db 319 gaccagagctctctgaagctgtgaagaaactctgctgagcgctatcacacttgggtgac 378
QY 366 atcaagataaagcggaggtgagtggaagatgatgctctatctctcctgcct 425
Db 379 atcaagataaagcggaggtgagtggaagatgatgctctatctctcctgcct 438
QY 426 gaaacctgttctaccattatagatcaaatgccctaaaaatgtagtaccctgtgaaaagga 485
Db 439 gaaacctgttctaccattatagatcaaatgccctaaaaatgtagtaccctgtgaaaagga 498
QY 486 caaataaagcaatgaatacat 506
Db 499 caaataaagcaatgaatacct 519
RESULT 7
AAX86241
ID AAX86241 standard; cDNA; 357 BP.
XX
AC AAX86241;
XX
DT 27-APR-1999 (first entry)
XX
DE EST clone AA246.
XX
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS Homo sapiens.
XX
PN W09845435-A2.
XX
PD 15-OCT-1998.
XX
PF 10-APR-1998; 98WO-US06954.
XX
PR 10-APR-1997; 97US-0835913.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;


```

XX US5939265-A.
XX PN
XX PD
XX PF 17-AUG-1999.
XX PR
XX PR 05-NOV-1997; 97US-0964725.
XX PR 05-NOV-1997; 97US-0964725.
XX PR 05-NOV-1996; 96US-0744211.
XX PR
XX PA (ABBO ) ABBOTT LAB.
XX PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
XX PI Kratochwill JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX PI WPI; 1999-468402/39.
XX DR
XX PT Polynucleotides useful for detecting, diagnosing and monitoring
XX PT diseases of the lung such as lung cancer
XX PS
XX PS Claim 1; Column 47; 36pp; English.
XX CC The 263 base pair sequence of a LU103 specific polynucleotide was
XX CC derived from clone 123531.
XX CC The polynucleotides and methods disclosed in the invention can be
XX CC useful for detecting, diagnosing, staging, monitoring or predicting
XX CC diseases and conditions of the lung, such as lung cancer.
XX CC The polynucleotides may be used to produce probes for use in
XX CC fluorescent in situ hybridization (FISH) technology to perform
XX CC chromosomal analysis and identify cancer specific alterations
XX CC such as deletions.
XX SQ Sequence 263 BP; 64 A; 57 C; 71 G; 71 T; 0 other;

Query Match 51.9%; Score 263; DB 20; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.2e-76;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 tgcactgccttctcatcaacaaagtgccttctgttgacaaagtggcacctttacc 195
Db 1 tgcactgccttctcatcaacaaagtgccttctgttgacaaagtggcacctttacc 60

QY 196 tctggacaacattcttccctttatggatccattaaagcttctctgaaaactctgggcat 255
Db 61 tctggacaacattcttccctttatggatccattaaagcttctctgaaaactctgggcat 120

QY 256 tctgttgagcaccttctggagggtgctaaagtgctgtaaatgagctgggaccagggc 315
Db 121 tctgttgagcaccttctggagggtgctaaagtgctgtaaatgagctgggaccagggc 180

QY 316 tctgaagctgtgaagaaactctggagcgctatcacacttgggtgacatcaagataa 375
Db 181 tctgaagctgtgaagaaactctggagcgctatcacacttgggtgacatcaagataa 240

QY 376 agagcgaggtggatggggatgg 398
Db 241 agagcgaggtggatggggatgg 263

RESULT 10
AAV38066
ID AAV38066 standard; cDNA; 269 BP.
XX
XX AAV38066;
XX
XX 14-SEP-1998 (first entry)
XX
XX Human lung tissue gene LU103 specific polynucleotide SEQ ID NO:1.
XX
XX Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.
XX
XX Homo sapiens.

```

```

XX WO9820143-A1.
XX PN
XX PD 14-MAY-1998.
XX PR
XX PR 05-NOV-1997; 97WO-US20680.
XX PR 05-NOV-1996; 96US-0744211.
XX PR
XX PA (ABBO ) ABBOTT LAB.
XX PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;
XX PI Kratochwill JD, Roberts-Rapp L, Russell JC, Stroupe SD;
XX PI WPI; 1998-286957/25.
XX DR
XX PT Lung tissue derived polynucleotide LU103 - useful to detect,
XX PT diagnose, stage, monitor, prognosis, prevent, treat or determine
XX PT pre-disposition to lung disease, e.g. lung cancer
XX PS
XX PS Claim 1; Page 66-67; 86pp; English.
XX CC The present sequence represents a polynucleotide specific for lung
XX CC tissue gene LU103. A method has been developed for detecting the
XX CC presence of a target LU103 polynucleotide in a test sample, comprising:
XX CC (a) contacting the sample with at least 1 LU103-specific polynucleotide,
XX CC and (b) detecting the target LU103 polynucleotide in the test sample,
XX CC where the LU103 polynucleotide has at least 50% identity to the 269,
XX CC 283, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to
XX CC AAV38070. The methods and products of the present invention may be used
XX CC to detect, diagnose, stage, monitor, prognosis, prevent, treat or
XX CC determine the predisposition diseases and conditions of the lung, e.g.
XX CC lung cancer.
XX SQ Sequence 269 BP; 62 A; 67 C; 48 G; 92 T; 0 other;

Query Match 51.9%; Score 263; DB 19; Length 269;
Best Local Similarity 100.0%; Pred. No. 3.2e-76;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gtggaacacactgctgtgtgatttctgagattttctgatttttaaacctcctgaaaaa 65
Db 7 gtggaacacactgctgtgtgatttctgagattttctgatttttaaacctcctgaaaaa 66

QY 66 tatccagataactgctgaagctgtaactatctctctgctgtgacacatcgccttt 125
Db 67 tatccagataactgctgaagctgtaactatctctctgctgtgacacatcgccttt 126

QY 126 gtattactctgctactgcttctctcatcaacaaagtgccttctctgttgacaaagtgg 185
Db 127 gtattactctgctactgcttctctcatcaacaaagtgccttctctgttgacaaagtgg 186

QY 186 cacctttacactctggacaacattcttccctttatggatccattaaagcttcttctgaaaa 245
Db 187 cacctttacactctggacaacattcttccctttatggatccattaaagcttcttctgaaaa 246

QY 246 ctctgggcatctctgtgagcac 268
Db 247 ctctgggcatctctgtgagcac 269

RESULT 11
AAV99421
ID AAV99421 standard; DNA; 269 BP.
XX
XX AAV99421;
XX
XX 19-OCT-1999 (first entry)
XX
XX LU103 specific polynucleotide isolated from clone 1235095.
XX
XX LU103; tumour; lung cancer; detection; FISH;

```

KW Fluorescent in situ hybridisation; ss.

XX Synthetic.
OS Homo sapiens.

XX US5939265-A.

XX 17-AUG-1999.

XX 05-NOV-1997; 97US-0964725.

XX 05-NOV-1997; 97US-0964725.

XX 05-NOV-1996; 96US-0744211.

XX (ABBO) ABBOTT LAB.

XX Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;

XX Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;

XX WPI; 1999-468402/39.

XX Polynucleotides useful for detecting, diagnosing and monitoring

XX diseases of the lung such as lung cancer

XX Claim 1; Column 47-48; 36pp; English.

XX The 269 base pair sequence of a LU103 specific polynucleotide was

XX derived from clone 1235095.

XX The polynucleotides and methods disclosed in the invention can be

XX useful for detecting, diagnosing, staging, monitoring or predicting

XX diseases and conditions of the lung, such as lung cancer.

XX The polynucleotides may be used to produce probes for use in

XX fluorescent in situ hybridization (FISH) technology to perform

XX chromosomal analysis and identify cancer specific alterations

XX such as deletions.

XX Sequence 269 BP; 62 A; 67 C; 48 G; 92 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

KW Fluorescent in situ hybridisation; ss.

XX Synthetic.
OS Homo sapiens.

XX US5939265-A.

XX 17-AUG-1999.

XX 05-NOV-1997; 97US-0964725.

XX 05-NOV-1997; 97US-0964725.

XX 05-NOV-1996; 96US-0744211.

XX (ABBO) ABBOTT LAB.

XX Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;

XX Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;

XX WPI; 1999-468402/39.

XX Polynucleotides useful for detecting, diagnosing and monitoring

XX diseases of the lung such as lung cancer

XX Claim 1; Column 47-48; 36pp; English.

XX The 269 base pair sequence of a LU103 specific polynucleotide was

XX derived from clone 1235095.

XX The polynucleotides and methods disclosed in the invention can be

XX useful for detecting, diagnosing, staging, monitoring or predicting

XX diseases and conditions of the lung, such as lung cancer.

XX The polynucleotides may be used to produce probes for use in

XX fluorescent in situ hybridization (FISH) technology to perform

XX chromosomal analysis and identify cancer specific alterations

XX such as deletions.

XX Sequence 269 BP; 62 A; 67 C; 48 G; 92 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Human; lung tissue gene; LU103; detection; lung cancer; diagnosis; ss.

Homo sapiens.

WO9820143-A1.

14-MAY-1998.

05-NOV-1997; 97WO-US20680.

05-NOV-1996; 96US-0744211.

(ABBO) ABBOTT LAB.

Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;

Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;

WPI; 1998-286957/25.

Lung tissue derived polynucleotide LU103 - useful to detect,

diagnose, stage, monitor, prognosis, prevent, treat or determine

pre-disposition to lung disease, e.g. lung cancer

Claim 1; Page 67; 86pp; English.

The present sequence represents a polynucleotide specific for lung

tissue gene LU103. A method has been developed for detecting the

presence of a target LU103 polynucleotide in a test sample, comprising:

(a) contacting the sample with at least 1 LU103-specific polynucleotide,

and (b) detecting the target LU103 polynucleotide in the test sample,

where the LU103 polynucleotide has at least 50% identity to the 269,

263, 225, 507 or 519 bp nucleic acid sequence given in AAV38066 to

AAV38070. The methods and products of the present invention may be used

to detect, diagnose, stage, monitor, prognosis, prevent, treat or

determine the predisposition diseases and conditions of the lung, e.g.

lung cancer.

Sequence 225 BP; 74 A; 41 C; 60 G; 50 T; 0 other;

Query Match 44.4%; Score 225; DB 19; Length 225;

Best Local Similarity 100.0%; Pred. No. 7.9e-64;

Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 283 aaggaagtgttaaatgagctgggaccagagcttctgaagctggaagaaactgctgga 342

Db 1 aaggaagtgttaaatgagctgggaccagagcttctgaagctggaagaaactgctgga 60

Qy 343 ggcctatcacacttgggtgacatcaagataaagagcgaggtggatggggaaga 402

Db 61 ggcctatcacacttgggtgacatcaagataaagagcgaggtggatggggaaga 120

Qy 403 tgatgctctatctctccctgctgaaacctgttctaccattatagatacaatgcctaa 462

Db 121 tgatgctctatctctccctgctgaaacctgttctaccattatagatacaatgcctaa 180

Qy 463 aatgtagtaccctgtaaaaggacaaataaagcaatgaatacatt 507

Db 181 aatgtagtaccctgtaaaaggacaaataaagcaatgaatacatt 225

RESULT 13

AAV38068

ID AAV38068 standard; cDNA; 225 BP.

XX AAV38068;

XX 19-OCT-1999 (first entry)

XX LU103 specific polynucleotide isolated from clone 1379417.

XX LU103; tumour; lung cancer; detection; FISH;

XX

XX

XX

XX

KW Fluorescent in situ hybridisation; ss.

XX Synthetic.

OS Homo sapiens.

XX US5939265-A.

PN 17-AUG-1999.

XX 05-NOV-1997; 97US-0964725.

XX 05-NOV-1997; 97US-0964725.

PR 05-NOV-1996; 96US-0744211.

XX (ABBO) ABBOTT LAB.

XX Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR;

PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;

XX WPI; 1999-468402/39.

XX Polynucleotides useful for detecting, diagnosing and monitoring

PT diseases of the lung such as lung cancer

XX Claim 1; Column 47; 36pp; English.

XX The 225 base pair sequence of a LU103 specific polynucleotide was

CC derived from clone 1379417.

CC The polynucleotides and methods disclosed in the invention can be

CC useful for detecting, diagnosing, staging, monitoring or predicting

CC diseases and conditions of the lung, such as lung cancer.

CC The polynucleotides may be used to produce probes for use in

CC fluorescent in situ hybridization (FISH) technology to perform

CC chromosomal analysis and identify cancer specific alterations

CC such as deletions.

XX Sequence 225 BP; 74 A; 41 C; 60 G; 50 T; 0 other;

SQ

Query Match

Best Local Similarity 44.4%; Score 225; DB 20; Length 225;

Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 aaggagtgtaaatgagctggaccagaggcttctgaagctgtgaagaaactgctgga 342

Db 1 aaggagtgtaaatgagctggaccagaggcttctgaagctgtgaagaaactgctgga 60

QY 343 ggcgctatcacacttggtgtgacatcaagaataaagcggagtggtggatggaaga 402

Db 61 ggcgctatcacacttggtgtgacatcaagaataaagcggagtggtggatggaaga 120

QY 403 tgatgctctatctccctgctggaacctgttctccaattatagatcaaatgccctaa 462

Db 121 tgatgctctatctccctgctggaacctgttctccaattatagatcaaatgccctaa 180

QY 463 aatgtagtgaccctgtaaaaggacaaataaagcaatgaatacatt 507

Db 181 aatgtagtgaccctgtaaaaggacaaataaagcaatgaatacatt 225

RESULT 14

AAT22859

ID AAT22859 standard; cDNA to mRNA; 60 BP.

XX AAT22859;

XX 29-AUG-1996 (first entry)

XX Human gene signature HUMGS04535.

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;

XX human; cloning; mapping; non-biased library; diagnosis; detection;

XX cell typing; abnormal cell function; ss.

XX

OS Homo sapiens.

XX WO9514772-A1.

PN 01-JUN-1995.

XX 11-NOV-1994; 94WO-JP01916.

XX 12-NOV-1993; 93JP-0355504.

XX (MATSU) MATSUBARA K.

PA (OKUBO) OKUBO K.

XX Matsubara K, Okubo K;

PI WPI; 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.

XX for diagnosis of abnormal cell function, by preparing cDNA that

XX reflects relative abundance of corresp. mRNA in specific human

XX tissues

XX Claim 1; Page 1229; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.

XX double-stranded DNA) which comprises one of the 7837 "GS" sequences

XX given in AAT19001-T26837 and which is able to hybridise to part of

XX human genomic DNA, cDNA or mRNA is claimed. The GS (gene Signature)

XX sequences were obtained from 3'-directed cDNA libraries prepared

XX from various human tissues; synthesis of cDNA was initiated from the

XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

XX untranslated sequence is unique to a particular mRNA species, almost

XX all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

XX is constructed so as to reflect accurately the relative abundance of

XX different mRNAs in the particular tissue from which it was derived.

XX The appearance frequency of a given GS in a cDNA library can be

XX determined (esp. using primers and probes derived from the GS

XX sequences) as a means of diagnosing abnormal cell function or for

XX recognising different cell types.

XX Sequence 60 BP; 29 A; 10 C; 11 G; 10 T; 0 other;

SQ

Query Match 11.4%; Score 58; DB 16; Length 60;

Best Local Similarity 100.0%; Pred. No. 1.7e-09;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 448 gatcaaatgccctaaatgtagtgaccctgaaaggacaaataaagcaatgaataca 505

Db 1 gatcaaatgccctaaatgtagtgaccctgaaaggacaaataaagcaatgaataca 58

RESULT 15

AAV54618

ID AAV54618 standard; cDNA; 225 BP.

XX AAV54618;

XX 30-OCT-1998 (first entry)

XX LU105 specific polynucleotide sequence from clone 1605935.

XX LU105; lung disease marker; immunoassay; lung disease; cancer;

XX blood; plasma; serum; ss.

XX Homo sapiens.

XX WO9833926-A1.

XX 06-AUG-1998.

XX 30-JAN-1998; 98WO-US01766.

PF

```

XX 31-JAN-1997; 97US-0791710.
XX (ABBO ) ABBOTT LAB.
XX Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
XX Gordon J, Granados, Hodges SC, Klass MR, Kratochvil JD;
XX Roberts-Rapp L, Russell JC, Stroupe SD;
XX WPI; 1998-437479/37.
XX
XX New nucleic acid for the lung disease marker LU105 - polypeptides,
XX antibodies and genes, used for diagnosis, prevention, treatment of
XX lung disease, specifically cancer
XX
XX Claim 11; Fig 1; 123pp; English.
XX
XX Sequences shown in AAV54616 to AAV54621 represent LU105 specific
XX polynucleotide sequences. These are used in the method of the invention
XX for detecting target LU105 nucleic acid. The method comprises treating a
XX sample with at least one LU105 specific nucleic acid, or its complement
XX which is at least 50 percent identical with the LU105 specific nucleic
XX acid sequences (AAV54616 to AAV54621). LU105 is a lung disease marker.
XX cells transformed with a recombinant expression system that contains
XX LU105 specific nucleic acid fragments, are used to express recombinant
XX LU105 polypeptides which are used to raise antibodies. The antibodies are
XX used to detect the LU105 antigen, and correspondingly this antigen is
XX used to detect specific antibodies, in usual immunoassays. The LU105
XX polypeptides and nucleic acid sequences are used for diagnosis, staging,
XX monitoring, prognosis, prevention, treatment and determination of
XX susceptibility to, lung disease, specifically cancer. The LU105
XX polypeptides are also used to screen for specific binding agents, useful
XX therapeutically. LU105 is a marker for lung disease (present at high
XX concentration, in altered form or in an unusual body compartment). LU105
XX can be detected in blood, plasma or serum in an inexpensive, non-invasive
XX test.
XX
XX Sequence 225 BP; 40 A; 76 C; 74 G; 35 T; 0 other;

```

```

Query Match      11.1%; Score 56.4; DB 19; Length 225;
Best Local Similarity 60.4%; Pred. NO. 1.le-08;
Matches 93; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 195 cttgtgacaacattctccctttatggatccattaaagcttctctgaaaactctggcca 254
      | ||| |||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 ccttggccaaacccctcgggaacctcaaccctgaagctcctctgtgagcctctggcca 67

QY 255 ttctgttgagacacctgtgtggagggtctgaaggaagtgtgtaaatgagctgggaccagagg 314
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 tccccgtgaaccacctcatagagggtctccacagaagtgtgtggtgagctgggtccccagg 127

QY 315 cttctgaagctgtgaagaactgctgtgagcgct 348
      | | | | | | | | | | | | | | | | | |
Db 128 ccgtggggccgtgaaggccctgaaggccctgct 161

```

Search completed: November 17, 2001, 14:50:26
Job time: 7154 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 12:47:26 ; Search time 4216.79 Seconds
(without alignments)
1859.744 Million cell updates/sec

Title: US-09-700-770-1
Perfect score: 507
Sequence: 1 ggcaagtgaaccactggct.....aataaagcaatgaatacatt 507

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_ba3:
4: gb_in1:
5: gb_in2:
6: gb_in3:
7: gb_om:
8: gb_ov:
9: gb_pat1:
10: gb_pat2:
11: gb_ph:
12: gb_pl1:
13: gb_pl2:
14: gb_pl3:
15: gb_pl4:
16: em_ba1:
17: em_ba2:
18: em_fun:
19: em_htgo_hum:
20: em_htgo_inv:
21: em_htgo_rod:
22: em_htg_hum1:
23: em_htg_hum2:
24: em_htg_hum3:
25: em_htg_hum4:
26: em_htg_hum5:
27: em_htg_hum6:
28: em_htg_hum7:
29: em_htg_hum8:
30: em_htg_inv1:
31: em_htg_inv2:
32: em_htg_other:
33: em_htg_rod:
34: em_hum1:
35: em_hum2:
36: em_hum3:
37: em_hum4:
38: em_hum5:
39: em_hum6:
40: em_hum7:
41: em_in:
42: em_om:
43: em_or:

44: em_ov:
45: em_pat:
46: em_ph:
47: em_pl:
48: em_ro:
49: em_sts:
50: em_sy:
51: em_un:
52: em_vi:
53: gb_sts1:
54: gb_sts2:
55: gb_sts3:
56: gb_sy:
57: gb_un:
58: gb_vil:
59: gb_vl2:
60: gb_htgl:
61: gb_htgl2:
62: gb_htgl3:
63: gb_htgl4:
64: gb_htgl5:
65: gb_htgl6:
66: gb_htgl7:
67: gb_htgl8:
68: gb_htgl9:
69: gb_htgl10:
70: gb_htgl11:
71: gb_htgl12:
72: gb_htgl13:
73: gb_htgl14:
74: gb_htgl15:
75: gb_htgl16:
76: gb_htgl17:
77: gb_htgl18:
78: gb_htgl19:
79: gb_htgl20:
80: gb_htgl21:
81: gb_htgl22:
82: gb_htgl23:
83: gb_htgl24:
84: gb_htgl25:
85: gb_pr1:
86: gb_pr2:
87: gb_pr3:
88: gb_pr4:
89: gb_pr5:
90: gb_pr6:
91: gb_pr7:
92: gb_pr8:
93: gb_pr9:
94: gb_rol:
95: gb_ro2:
96: gb_in4:
97: gb_pr10:
98: em_ba3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	205.4	40.5	166539	62	AC011402
2	192.8	38.0	160949	62	AC011352
3	157	31.0	207747	62	AC011334
C 4	57.8	11.4	96482	67	AC022095
C 5	57.8	11.4	168347	69	AC025336
6	56.4	11.1	471	97	HUMZB52D10
7	38.2	7.5	148391	71	AC034174
8	38	7.5	7218	10	I66494

Sequence 14

*****		Summary Statistics	
Consensus quality: 73731 bases at least Q40		Consensus quality: 84145 bases at least Q30	
Consensus quality: 84145 bases at least Q30		Consensus quality: 88496 bases at least Q20	
Estimated insert size: 131000; pulse field gel estimation		Estimated insert size: 93182; sum-of-contigs estimation	
Quality coverage: 4.03 in Q20 bases; pulse field gel estimation		Quality coverage: 5.54 in Q20 bases; sum-of-contigs estimation.	
* NOTE: This is a 'working draft' sequence. It currently		* consists of 14 contigs. The true order of the pieces	
* is not known and their order in this sequence record is		* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.		* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will		* be preserved.	
1		2069: contig of 2069 bp in length	
2070		2169: gap of unknown length	
2170		4188: contig of 2019 bp in length	
4189		4288: gap of unknown length	
4289		6338: contig of 2050 bp in length	
6439		6439: gap of unknown length	
6439		8789: contig of 2351 bp in length	
8790		8889: gap of unknown length	
8890		11124: contig of 2235 bp in length	
11125		11224: gap of unknown length	
11225		13472: contig of 2248 bp in length	
13473		13572: gap of unknown length	
13573		16709: contig of 3137 bp in length	
16710		16809: gap of unknown length	
16810		20134: contig of 3325 bp in length	
20135		20234: gap of unknown length	
20235		23845: contig of 3611 bp in length	
23846		23945: gap of unknown length	
23946		28527: contig of 4582 bp in length	
28528		28527: gap of unknown length	
34284		34284: contig of 5657 bp in length	
34285		45172: contig of 10788 bp in length	
45173		45272: gap of unknown length	
45273		56033: contig of 10761 bp in length	
56034		56133: gap of unknown length	
56134		96482: contig of 40349 bp in length.	
FEATURES		Location/Qualifiers	
1..96482		/organism="Homo sapiens"	
/db_xref="taxon:9606"		/chromosome="5"	
/clone="CTB-36B8"		/clone.lib="Caltech human BAC library B"	
BASE COUNT		21454 a 25145 c 26218 g 22348 t 1317 others	
ORIGIN			
Query Match		11.4%; Score 57.8; DB 67; Length 96482;	
Best Local Similarity		59.4%; Pred. No. 2.4e-05;	
Matches		98; Conservative 0; Mismatches 67; Indels 0; Gaps	
QY		195 ctctggacaacattcttcctttatggatccattaaagcttcttctgaaacctgggcca 254	
Db		18605 CCCTGGCCCAACCCCTCGGACCCCTCAACCGCTGAAGCTCTCTGCTGAGCAGCTGGGCA 18546	
QY		255 ttctgttgagcacctgtgaggggctaaagagtggtgtaaatgactgggacacagg 314	
Db		18545 TCCCGGTGAACCACTCATAGAGGGCTCCACAGAGTGTGGCTGGTGGTCCCCCAGG 18486	
QY		315 ctctggaagctgtgaagaactgctggagcgctatcacacttgg 359	
Db		18485 CCGTGGGGCGCGTGAAGGCCCTGAAGGCCCTGCTGTGTAAGTGGG 18441	
RESULT		5	

AC025336/c

LOCUS

DEFINITION

AC025336

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC025336 168347 bp DNA HTG 25-MAR-2000
Homo sapiens chromosome 5 clone RP11-451H23 map 5, WORKING DRAFT
SEQUENCE, 32 unordered pieces.
AC025336
AC025336.2 GI:7328761
HTG: HTGS-PHASE1; HTGS-DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 168347)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 5, clone RP11-451H23
Unpublished
2 (bases 1 to 168347)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckhagter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehocsky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 25, 2000 this sequence version replaced gi:7210017.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smith,A.F.A. & Green, P. (1996-1997)
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6686
Center clone name: 451_H_23
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 150422 bases at least Q40
Consensus quality: 159524 bases at least Q30
Consensus quality: 163013 bases at least Q20
Insert size: 165247; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1389: contig of 1389 bp in length
* 1390 1489: gap of 100 bp
* 1490 3130: contig of 1641 bp in length

3131 3230: gap of 100 bp
3231 4942: contig of 1712 bp in length
4943 5042: gap of 100 bp
5043 6981: contig of 1939 bp in length
6982 7081: gap of 100 bp
7082 8708: contig of 1627 bp in length
8709 8808: gap of 100 bp
8809 10286: contig of 1478 bp in length
10287 10386: gap of 100 bp
10387 12212: contig of 1826 bp in length
12213 12312: gap of 100 bp
12313 14658: contig of 2346 bp in length
14659 14758: gap of 100 bp
14759 17941: contig of 3183 bp in length
17942 18041: gap of 100 bp
18042 21297: contig of 3256 bp in length
21298 21397: gap of 100 bp
21398 24992: contig of 3595 bp in length
24993 25092: gap of 100 bp
25093 27768: contig of 2676 bp in length
27769 27868: gap of 100 bp
27869 31188: contig of 3320 bp in length
31189 31288: gap of 100 bp
31289 33714: contig of 2426 bp in length
33715 33814: gap of 100 bp
33815 37277: contig of 3463 bp in length
37278 37377: gap of 100 bp
37378 42302: contig of 4925 bp in length
42303 42402: gap of 100 bp
42403 47816: contig of 5414 bp in length
47817 47916: gap of 100 bp
47917 52586: contig of 4670 bp in length
52587 52686: gap of 100 bp
52687 56567: contig of 3881 bp in length
56568 61557: contig of 4890 bp in length
61558 61657: gap of 100 bp
61658 66724: contig of 5067 bp in length
66725 66824: gap of 100 bp
66825 71568: contig of 4744 bp in length
71569 71668: gap of 100 bp
71669 76578: contig of 4910 bp in length
76579 76678: gap of 100 bp
76679 83312: contig of 6634 bp in length
83313 83412: gap of 100 bp
83413 90053: contig of 6641 bp in length
90054 90153: gap of 100 bp
90154 99426: contig of 9273 bp in length
99427 99526: gap of 100 bp
99527 108015: contig of 8489 bp in length
108016 108115: gap of 100 bp
108116 118144: contig of 10029 bp in length
118145 118244: gap of 100 bp
118245 130468: contig of 12224 bp in length
130469 130568: gap of 100 bp
130569 142239: contig of 11671 bp in length
142240 142339: gap of 100 bp
142340 157135: contig of 14796 bp in length
157136 157235: gap of 100 bp
157236 168347: contig of 11112 bp in length.
Location/Qualifiers
1. .168347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5"
/clone_lib="RP11-451H23"
/clone="RP11-451H23"
1. .1389
/note="assembly_fragment"
1490. .3130
/note="assembly_fragment"
3231. .4942
misc_feature
misc_feature
misc_feature

FEATURES
source

RESULT 7

AC034174
LOCUS
DEFINITION
Homo sapiens chromosome 8 clone RP11-713C9, WORKING DRAFT SEQUENCE,
34 unordered pieces.
AC034174
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 148391)
Waterston,R.H.
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (04-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT

On May 26, 2000 this sequence version replaced gi:7579889.

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H_NH0713C09

----- Summary Statistics -----

Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%

Chemistry: Dye-primer ET; 100% of reads

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 130554 bases at least Q40

Consensus quality: 135358 bases at least Q30

Consensus quality: 137816 bases at least Q20

Insert size: 144000; agarose-fp

Insert size: 145091; sum-of-contigs

Quality coverage: 3.13 in Q20 bases; agarose-fp

Quality coverage: 3.22 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1305: contig of 1305 bp in length

* 1306 1405: gap of unknown length

* 1406 2475: contig of 1070 bp in length

* 2476 2575: gap of unknown length

* 2576 4079: contig of 1504 bp in length

* 4080 4179: gap of unknown length

* 4180 5373: contig of 1194 bp in length

* 5374 5473: gap of unknown length

* 5474 6612: contig of 1139 bp in length

* 6613 6712: gap of unknown length

* 6713 8233: contig of 1521 bp in length

* 8234 8334: gap of unknown length

* 8334 9870: contig of 1537 bp in length

* 9871 9971: gap of unknown length

* 9971 11307: contig of 1336 bp in length

* 11307 11406: gap of unknown length

* 11407 13135: contig of 1729 bp in length

* 13136 13235: gap of unknown length

* 13236 15265: contig of 2030 bp in length

* 15266 15365: gap of unknown length

* 15366 17713: contig of 2348 bp in length

* 17714 17813: gap of unknown length

* 17814 20069: contig of 2256 bp in length

* 20070 20169: gap of unknown length

* 20170 23481: contig of 3292 bp in length

* 23482 23561: gap of unknown length

* 23562 26701: contig of 3139 bp in length

* 26702 26800: gap of unknown length

* 26801 29564: contig of 2764 bp in length

* 29565 29664: gap of unknown length

* 29665 33722: contig of 4058 bp in length

* 33723 33822: gap of unknown length

* 33823 38003: contig of 4181 bp in length

* 38004 38103: gap of unknown length

* 38104 42998: contig of 4895 bp in length

* 42999 43098: gap of unknown length

* 43099 47470: contig of 4372 bp in length

* 47471 47570: gap of unknown length

* 47571 51371: contig of 3801 bp in length

* 51372 51471: gap of unknown length

* 51472 55769: contig of 4298 bp in length

* 55770 55869: gap of unknown length

* 55870 60125: contig of 4156 bp in length

* 60126 60226: gap of unknown length

* 60227 65249: contig of 5124 bp in length

* 65250 65349: gap of unknown length

* 65350 72542: contig of 7193 bp in length

* 72543 72642: gap of unknown length

* 72643 79514: contig of 6872 bp in length

* 79515 79614: gap of unknown length

* 79615 86014: contig of 6399 bp in length

* 86015 86114: contig of 6621 bp in length

* 86115 92735: gap of unknown length

* 92736 92834: gap of unknown length

* 92835 100432: contig of 7597 bp in length

* 100433 100531: gap of unknown length

* 100532 106203: contig of 5672 bp in length

* 106204 106304: gap of unknown length

* 106305 112770: contig of 6467 bp in length

* 112771 112871: gap of unknown length

* 112872 120318: contig of 7448 bp in length

* 120319 120418: gap of unknown length

* 120419 128797: contig of 8379 bp in length

* 128798 128897: gap of unknown length

* 128898 136590: contig of 7692 bp in length

* 136591 136690: gap of unknown length

* 136691 148391: contig of 11702 bp in length.

FEATURES

source

1. 148391
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="RP11-713C9"

BASE COUNT 48821 a 26107 c 26007 g 44143 t 3313 others
ORIGIN

Query Match 7.5%; Score 38.2; DB 71; Length 148391;
Best Local Similarity 52.9%; Pred. NO. 8.8;
Matches 82; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 221 gatccattaaagcttctctgaaacctctggcattctctgttgagacacctgtggagggg 280
Db 69720 GAATTATTGCAATCATGATAAACTTTGGTGAGTGAGGAGTTCTTCATTATGGAGGAG 69779

QY 281 ctaagaagtgtgtataatgagctggaccagagccttctgaagctgtgaagaactgctg 340
Db 69780 CAAAGAAAGTGGTTCTTCAGATGGAACCTACTCTTATGAAATGCTGTGAATATGTTT 69839

QY 341 gaggcgtatcacacttgggtgtgacatcaagataa 375
Db 69840 GAAATGATACAAATATTATTAGATATATACATTAA 69874

RESULT	8
LOCUS	I66494
DEFINITION	7218 bp DNA
ACCESSION	Sequence 14 from patent US 5670367.
VERSION	I66494.1 GI:2724471
KEYWORDS	
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE	Recombinant fowlpox virus
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;
FEATURES	Location/Qualifiers
source	1..7218
BASE COUNT	/organism="unknown"
ORIGIN	144 a 1491 c 1486 g 1929 t 368 others
Query Match	7.5%; Score 38; DB 10; Length 7218;
Best Local Similarity	4.3%; Pred.No.6.2;
Matches	11; Conservative 144; Mismatches 99; Indels 0; Gaps 0;
QY	18 gcttggtgatttctagattttctgaatttttaactcctcgaaaatatccagataa 77
Db	1048 GTCTGAGGGAGCTTCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1107
QY	.78 ctgctcatgaagctggtgaactatcttcctgcgtgtagaccatcacgtttggcaacatccccagataa 77
Db	1108 YY 137
QY	138 ctactgccttctcatcaacaagtgcacctctcgtgtgacaagttggcaaccttacctc 197
Db	1168 YY 1227
QY	198 tggacaacattcttcctttatgatcatcaaagctctctctgaaactctggcattt 257
Db	1228 YY 1287
QY	258 ctggtgagcaacctt 271
Db	1288 YYYYYYYYYYYYYY 1301
RESULT	9.
LOCUS	AL356463
DEFINITION	AL356463 154268 bp DNA HTG 23-JAN-2001
ACCESSION	Homo sapiens chromosome 1 clone RP11-100E13, *** SEQUENCING IN
VERSION	PROGRESS ***, 17 unordered pieces.
KEYWORDS	AL356463.4 GI:10039903
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT.
ORGANISM	human.
REFERENCE.	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	Plumb,B.
COMMENT	Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk On Sep 9, 2000 this sequence version replaced gi:9650562. ----- Ganome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: ba100E13 ----- Summary Statistics

```

/note="assembly_fragment:00637
fragment_chain:1"
misc_feature
40604..44208
/note="assembly_fragment:00475
fragment_chain:2"
misc_feature
44309..47023
/note="assembly_fragment:00804
fragment_chain:2"
misc_feature
47124..51832
/note="assembly_fragment:00574
fragment_chain:3"
misc_feature
51933..55546
/note="assembly_fragment:00320
fragment_chain:3"
misc_feature
55647..64021
/note="assembly_fragment:01111
fragment_chain:4"
misc_feature
64122..68879
/note="assembly_fragment:01077
fragment_chain:4"
misc_feature
68980..86316
/note="assembly_fragment:01193
fragment_chain:5"
misc_feature
86417..132139
/note="assembly_fragment:00263
fragment_chain:5"
misc_feature
132240..137720
/note="assembly_fragment:00001"
misc_feature
137821..141195
/note="assembly_fragment:00435"
misc_feature
141296..147431
/note="assembly_fragment:00657"
misc_feature
147532..154268
/note="assembly_fragment:00934"

```

```

BASE COUNT 42324 a 32439 c 34181 g 43718 t 1606 others
ORIGIN

```

```

Query Match 7.4%; Score 37.6; DB 80; Length 154268;
Best Local Similarity 54.3%; Pred. No. 13;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 98 atctctcgtgtgaccatcagccttctgttagtctactgtcgtcctctctcctcaac 157
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116404 ATCATGTTGTGACGTGGCCCTGCTCTGTATCTTCTTAGCTTTTCCCTTTAAATCGT 116463

QY 158 aaagtgcctctcctgttgacacagttggcactttacgtctgactgctcctcctcaac 217
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116464 TGATTTCTCTTATGTCACAGATGCGTGTACTCTCCAGGCATCAGGTCGTGCATT 116523

QY 218 atggatccattaaagcttct 237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116524 CAGCAAAAGAAATATAT 116543

RESULT 10
AC026235/c AC026235 157122 bp DNA HTG 22-FEB-2001
LOCUS Homo sapiens chromosome 1 clone RP11-285A20, WORKING DRAFT
DEFINITION SEQUENCE, 2 unordered pieces.
ACCESSION AC026235
VERSION AC026235.12 GI:12313761
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157122)
REFERENCE
AUTHORS Abola, A.P., Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D.,
Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hymen, R.,
Mao, J., Lam, B., Marathe, R., Miranda, M., Morehouse, A.J., Nguyen, M.,
Oefner, P., Palm, C.J., Ramirez, D., Southwick, A.M., Wilhelm, J.,
Yu, S. and Davis, R.W.

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 157122)
AUTHORS Abola, A.P., Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D.,
Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hymen, R.,
Mao, J., Marathe, R., Morehouse, A.J., Oefner, P., Palm, C.J.,
Ramirez, D., Wilhelm, J., Yu, S. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT On Jan 19, 2001 this sequence version replaced gi:12280739.
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center

```

```

Center code: SDSTD
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum-info@sequence.stanford.edu
----- Project Information
Center project name: 867
Center clone name: RP11-285A20
----- Summary Statistics
Sequencing Vector: M13mp18; X02513
Chemistry: Dye-primer; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 156451 bases at least Q40
Consensus quality: 156776 bases at least Q30
Consensus quality: 156936 bases at least Q20
Insert size: 162013; agarose-fp
Insert size: 157022; sum-of-contigs
Quality coverage: 10.7x in Q20 bases; agarose-fp
Quality coverage: 11.0x in Q20 bases; sum-of-contigs.

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9109: contig of 9109 bp in length
* 9110 9209: gap of unknown length
* 9210 157122: contig of 147913 bp in length.
FEATURES
Location/Qualifiers
source
1..157122
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-285A20"
/clone_lib="RPC1 human BAC library 11"
misc_feature
1..9109
/note="assembly_name:Contig15
clone_end:SP6"
misc_feature
9210..157122
/note="assembly_name:Contig16
clone_end:T7"

```

```

BASE COUNT 45704 a 35408 c 33391 g 42519 t 100 others
ORIGIN

```

```

Query Match 7.4%; Score 37.6; DB 69; Length 157122;
Best Local Similarity 54.3%; Pred. No. 13;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 98 atctctcgtgtgaccatcagccttctgttagtctactgctcctcctcctcaac 157
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59739 ATCATGTTGTGACGTGGCCCTGCTCTGTATCTTCTTAGCTTTTCCCTTTAAATCGT 59680

QY 158 aaagtgcctctcctgttgacacagttggcactttacgtctgactgctcctcctccttt 217
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59679 TGATTTCTCTTATGTCACAGATGCGTGTACTCTCCAGGCATCAGGTCGTGCATT 59620

QY 218 atggatccattaaagcttct 237

```

[illegible]

```

* 1 5772: contig of 5772 bp in length
* 5773 5872: gap of unknown length
* 10133 10133: contig of 4261 bp in length
* 10134 10586: gap of unknown length
* 10234 15586: contig of 5353 bp in length
* 15387 15686: gap of unknown length
* 15687 20246: contig of 4560 bp in length
* 20247 26852: contig of 6506 bp in length
* 26853 36105: gap of unknown length
* 26953 36205: contig of 9153 bp in length
* 36106 44497: gap of unknown length
* 44498 44597: contig of 8292 bp in length
* 44598 53137: gap of unknown length
* 53138 53237: contig of 8540 bp in length
* 53238 62667: gap of unknown length
* 62668 62767: contig of 9430 bp in length
* 62768 71276: gap of unknown length
* 71277 81084: contig of 8509 bp in length
* 81085 81184: gap of unknown length
* 81185 95747: contig of 14563 bp in length
* 95748 95847: gap of unknown length
* 95848 112911: contig of 17064 bp in length
* 112912 132279: gap of unknown length
* 132280 132379: contig of 19268 bp in length
* 132380 159697: gap of unknown length
* 159698 215455: contig of 27318 bp in length
* 159798 215455: contig of 55658 bp in length.

```

FEATURES

```

source
1. .215455
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="12"
/clone="RP23-366H13"
/clone_lib="RPCI mouse BAC library 23"
1. .5772
/note="assembly_fragment"
5873. .10133
/note="assembly_fragment"
/clone_end:SP6
vector_side:right
10234. .15586
/note="assembly_fragment"
15687. .20246
/note="assembly_fragment"
/clone_end:T7
vector_side:right
20347. .26852
/note="assembly_fragment"
26953. .36105
/note="assembly_fragment"
36206. .44497
/note="assembly_fragment"
44598. .53137
/note="assembly_fragment"
53238. .62667
/note="assembly_fragment"
62768. .71276
/note="assembly_fragment"
71377. .81084
/note="assembly_fragment"
81185. .95747
/note="assembly_fragment"
95848. .112911
/note="assembly_fragment"
113012. .132279
/note="assembly_fragment"
132380. .159697
/note="assembly_fragment"

```

```

misc_feature 159798..215455
/note="assembly_fragment"
BASE COUNT 68411 a 40237 c 39791 g 65506 t 1510 others
ORIGIN

Query Match 7.4%; Score 37.6; DB 76; Length 215455;
Best Local Similarity 49.0%; Pred. No. 14;
Matches 100; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 35 agattttctgatttttaaaactcctcgaataatccccagataaactgctgaagctggtta 94
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142794 ACATTACTCATTTAGTTGACCCAGAGAGAGGAGTGAACACACAGCCTTCATAAGAAGATT 142735
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 95 actatctctcgtcgtggagaccatcagcctttgttagtctactcgtcactgctctctcctc 154
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142734 ATTTTGTGTTAAACCATGCTGCATCTTTTGTGTAAACACAGATTAAACAATCCCATG 142675
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 155 acaaaagtgcctctctctgttgacaagtggcacctcttaccctctggacaacattctctcc 214
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142674 TATCTTGCTGCTTTTCTCTCTATAGGAAGCATCTATATATTTAAACATTATTGATCCC 142615
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 215 ttatgatccattaaagctctctt 238
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142614 TATATCTATCCGTTAAATGTTTCTT 142591
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AC006508 185286 bp DNA ROD 29-AUG-2000
LOCUS Mus musculus Yp BAC GSMB-187H15 (Genome Systems Mouse BAC Library)
DEFINITION complete sequence.
ACCESSION AC006508 AC006401
VERSION AC006508.2 GI:4454569
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 185286)
AUTHORS Muzny,D., Chen,Z., Culpepper,P., Ding,Y., Dugan,S.P., Durbin,K.J.,
Esparanza,J., Forcum,J., Garcia,C., Gorrell,H., Gorrell,L.L.,
Hernandez,J., Huang,W.J., Jackson,L.E., Kondejewski,N., Leal,B.,
Lichtarge,O., Liu,W., Logan,O., Lu,J., Martinez,C., Oswal,G.,
Pampell,L.R., Parish,B.J., Perez,L.M., Rashid,N.D., Rives,C.M.,
Scherer,S.E., Shen,H., Simon,M.L., Ty,T.M., Vo,Q.K., Wei,Y.,
Williamson,A.L., Worley,K., Zhou,X., Bishop,C.E. and Gibbs,R.A.
Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185286)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-1999) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 185286)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 185286)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 29, 2000 this sequence version replaced gi:4406719
gi:4225890.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the

```

entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids. Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES
source

Location/Qualifiers
1. .185286

/organism="Mus musculus"
/db_xref="taxon:10090"

/chromosome="Yp"
/clone="GSM8-187H15"

join(386. .486,2580. .2641,4014. .4062,6524. .6653,7385. .7540,7967. .8066,17122. .17257,19778. .19864,20273. .20371,20447. .20607,20993. .21137,21247. .21391,21757. .21938,22008. .22125,23018. .23171,23267. .23397,23600. .25372)

/gene="Mus musculus mRNA for DBY RNA helicase AJ007376"

complement(1086. .1201)
/rpt_family="B1_MM"

complement(1523. .1645)
/rpt_family="B3"

complement(2224. .2438)
/rpt_family="B3"

4289. .4421
/rpt_family="B1-F"

4617. .4785
/rpt_family="B3"

complement(5462. .5644)
/rpt_family="B2"

complement(6735. .7044)
/rpt_family="MT2B"

complement(7826. .7852)
/rpt_family="AT-rich"

complement(9772. .12427)
/rpt_family="IAP"

complement(12822. .13010)
/rpt_family="HERVK"

complement(13168. .14064)
/rpt_family="IAP"

complement(15303. .15357)
/rpt_family="(CCAA)n"

complement(15425. .15581)
/rpt_family="B2"

complement(15595. .16288)
/rpt_family="Lx9"

repeat_region complement(16638. .16706)
/rpt_family="B4A"
repeat_region complement(16695. .16871)
/rpt_family="B3"
repeat_region complement(17360. .17412)
/rpt_family="ORR1B"
repeat_region complement(17554. .17726)
/rpt_family="ORR1B"
17793. .17938
/rpt_family="B1_MM"
17960. .18000
/rpt_family="(GA)n"
18582. .18691
/rpt_family="(GAAA)n"
repeat_region complement(18696. .18883)
/rpt_family="ORR1C"
19321. .19403
/rpt_family="B1-F"
repeat_region complement(19518. .19589)
/rpt_family="PBID9"
20065. .20208
/rpt_family="B3"
repeat_region complement(21535. .21619)
/rpt_family="B4A"
22548. .22593
/rpt_family="B1-F"
22711. .22846
/rpt_family="B1-F"
repeat_region complement(23403. .23451)
/rpt_family="(CAAA)n"
24630. .24700
/rpt_family="(TAAAA)n"
25239. .25387
/rpt_family="B1_MM"
26303. .26449
/rpt_family="B1_MM"
26534. .26705
/rpt_family="B2"
repeat_region complement(26852. .26972)
/rpt_family="B1_MM"
repeat_region complement(27305. .27426)
/rpt_family="B1_MM"
27452. .27653
/rpt_family="Lx8"
28137. .28342
/rpt_family="B2"
30010. .30107
/rpt_family="MYSERV"
repeat_region complement(30147. .30346)
/rpt_family="B2"
30955. .31375
/rpt_family="RMR17C"
repeat_region complement(31444. .31926)
/rpt_family="RMR15"
repeat_region complement(31927. .32058)
/rpt_family="ORR1B"
repeat_region complement(32065. .32178)
/rpt_family="RMR15"
repeat_region complement(32885. .33207)
/rpt_family="RLTRETN_MM"
repeat_region complement(34002. .34115)
/rpt_family="MT2B"
34450. .34552
/rpt_family="B2"
35167. .35356
/rpt_family="B2"
35526. .35626
/rpt_family="RSINEL"
repeat_region 35920. .35996
/rpt_family="B3A"
35997. .36095
/rpt_family="RLTRI3B"
36107. .36953
repeat_region

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 13:33:57 ; Search time 2570.26 Seconds
(without alignments)
7576.229 Million cell updates/sec

Title: US-09-700-770-3
Perfect score: 2060
Sequence: 1 cttgagagctctcaataact.....ttccattgaaaaa 2060

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_est3.*
- 4: gb_est4.*
- 5: gb_est5.*
- 6: gb_est6.*
- 7: gb_est7.*
- 8: gb_est8.*
- 9: gb_est9.*
- 10: gb_est10.*
- 11: gb_est11.*
- 12: gb_est12.*
- 13: gb_est13.*
- 14: gb_est14.*
- 15: gb_est15.*
- 16: gb_est16.*
- 17: gb_est17.*
- 18: gb_est18.*
- 19: gb_est19.*
- 20: gb_est20.*
- 21: gb_est21.*
- 22: gb_est22.*
- 23: gb_est23.*
- 24: gb_est24.*
- 25: gb_est25.*
- 26: gb_est26.*
- 27: gb_est27.*
- 28: gb_est28.*
- 29: gb_est29.*
- 30: gb_est30.*
- 31: gb_est31.*
- 32: gb_est32.*
- 33: gb_est33.*
- 34: gb_est34.*
- 35: gb_est35.*
- 36: gb_est36.*
- 37: gb_est37.*
- 38: gb_est38.*
- 39: gb_est39.*
- 40: gb_est40.*
- 41: gb_est41.*
- 42: gb_est42.*
- 43: gb_est43.*
- 44: gb_est44.*
- 45: gb_est45.*
- 46: gb_est46.*
- 47: gb_est47.*

- 44: em_esthum10.*
- 45: em_esthum11.*
- 46: em_esthum12.*
- 47: em_esthum13.*
- 48: em_esthum14.*
- 49: em_esthum15.*
- 50: em_esthum16.*
- 51: em_esthum17.*
- 52: em_esthum18.*
- 53: em_esthum19.*
- 54: em_esthum20.*
- 55: em_esthum21.*
- 56: em_esthum22.*
- 57: em_esthum23.*
- 58: em_esthum24.*
- 59: em_esthum25.*
- 60: em_esthum26.*
- 61: em_esthum27.*
- 62: em_esthum28.*
- 63: em_estin1.*
- 64: em_estin2.*
- 65: em_estin3.*
- 66: em_estin4.*
- 67: em_estin5.*
- 68: em_estom1.*
- 69: em_estov1.*
- 70: em_estov2.*
- 71: em_estpl1.*
- 72: em_estpl2.*
- 73: em_estpl3.*
- 74: em_estpl4.*
- 75: em_estpl5.*
- 76: em_estpl6.*
- 77: em_estpl7.*
- 78: em_estpl8.*
- 79: em_estpl9.*
- 80: em_estpl10.*
- 81: em_estro1.*
- 82: em_estro2.*
- 83: em_estro3.*
- 84: em_estro4.*
- 85: em_estro5.*
- 86: em_estro6.*
- 87: em_estro7.*
- 88: em_estro8.*
- 89: em_estro9.*
- 90: em_estro10.*
- 91: em_estro11.*
- 92: em_estro12.*
- 93: em_estro13.*
- 94: em_estro14.*
- 95: em_estro15.*
- 96: em_estro16.*
- 97: em_estro17.*
- 98: em_estro18.*
- 99: em_estro19.*
- 100: em_estro20.*
- 101: gb_est25.*
- 102: gb_est26.*
- 103: gb_est27.*
- 104: gb_est28.*
- 105: gb_est29.*
- 106: gb_est30.*
- 107: gb_est31.*
- 108: gb_est32.*
- 109: gb_est33.*
- 110: gb_est34.*
- 111: gb_est35.*
- 112: gb_est36.*
- 113: gb_est37.*
- 114: gb_est38.*
- 115: gb_est39.*
- 116: gb_est40.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred: No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	789	38.3	1331	192	AK002520	Mus musculus
2	696	33.8	878	155	BG540257	BG540257 602569148
3	662	32.1	901	155	BG541671	BG541671 602571247
4	648.2	31.5	729	166	A1129970	A1129970 qc50a08.x
5	617	30.0	849	153	BG397752	BG397752 602438820
6	613.4	29.8	965	155	BG542436	BG542436 602569511
7	611	29.7	804	110	AW025398	AW025398 wu96g01.x
8	609.4	29.6	770	175	BG283753	BG283753 602408045
9	605.2	29.4	612	103	A1924523	A1924523 w61b10.x
10	587.4	28.5	933	106	AL545162	AL545162 AL545162
11	572.4	27.8	762	24	A1763426	A1763426 wh92c02.x
12	562.2	27.3	587	170	BF869544	BF869544 IL3-ET011
13	558.2	27.1	880	141	BE992689	BE992689 601433907
14	552.4	26.8	813	138	BE622333	BE622333 601441119
15	550	26.7	837	111	AA744560	AA744560 nv79d11.s
16	549	26.7	1752	192	AK010310	AK010310 Mus muscu
17	546	26.5	578	149	BF513122	BF513122 UT-H-BW1-
18	532.6	25.9	701	23	A1679000	A1679000 tu60f11.x
19	530.6	25.8	1081	106	AL568646	AL568646 AL568646
20	525.2	25.5	793	153	BG428427	BG428427 602494930
21	514	25.0	514	168	BF724429	BF724429 bx04g09.Y
22	511.4	24.8	625	146	BF238238	BF238238 601811987
23	506.8	24.6	686	138	BE675393	BE675393 7f08c08.x
24	498.4	24.2	924	153	BG403561	BG403561 602419268
25	490.8	23.8	685	138	BE676409	BE676409 7f29b12.x
26	490.4	23.8	776	106	AL540913	AL540913 AL540913
27	485.8	23.6	653	113	AW268846	AW268846 xv48b06.x
28	485.6	23.6	638	102	A1806766	A1806766 wf15a04.x
29	485.2	23.6	639	138	BE674668	BE674668 7e93h11.x
30	483	23.4	986	106	AL555946	AL555946 AL555946
31	482.4	23.4	495	7	AA405967	AA405967 zu66f02.r
32	478.8	23.2	648	19	A1382062	A1382062 te68b01.x
33	477	23.2	638	24	A1760722	A1760722 w108d10.x
34	474.4	23.0	643	116	AW474169	AW474169 xy11h08.x
35	468.6	22.7	532	22	A1568939	A1568939 tg71f10.x
36	464	22.5	630	102	A1832066	A1832066 wj99h10.x
37	463.4	22.5	766	104	AL596435	AL596435 ul77b09.y
38	463.4	22.5	1022	106	AL549874	AL549874 AL549874
39	455.6	22.1	968	106	AL546499	AL546499 AL546499
40	455.2	22.1	630	17	A1221418	A1221418 qg92d06.x
41	454.4	22.1	727	11	AA769575	AA769575 nz42a01.s
42	454	22.0	628	21	A1523866	A1523866 tg97c03.x
43	452.4	22.0	769	104	A1987619	A1987619 ul86e10.y
44	449.2	21.8	786	155	BG536455	BG536455 602564554
45	443.4	21.5	492	17	A1191525	A1191525 qe44d11.x

ALIGNMENTS

RESULT	1
ACCESSION	AK002520
LOCUS	Mus musculus adult male kidney cDNA, HTC 08-FEB-2001
DEFINITION	library, clone:0610011C03, full insert sequence.
VERSION	AK002520
KEYWORDS	CAP trapper.
SOURCE	Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library
ORGANISM	clone:0610011C03. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS	1 (sites) Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning

JOURNAL
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORSTITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORSTITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORSTITLE
JOURNAL

COMMENT

FEATURES
source

CDS

Methods Enzymol. 303, 19-44 (1999)

2 (sites)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374

3 (sites)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Okazaki, Y., Muramatsu, M., Inoue, Y. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913

4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 1331)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiradaka, T., Hori, F., Imoto, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Soabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGCGCGCGACGAGTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGAGTCCAGAGCTCAATTAATTAATTAACCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

Location/Qualifiers
1. .1331
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:MGI:109365"
/db_xref="MGD:MGI:1907153"
/clone="0610011C03"
/sex="male"
/tissue-type="kidney"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
19. .1278

```
/note="putative"
/codon_start=1
/protein_id="BAB22158.1"
/db_xref="GI:12832561"
/translation="MSPILLLLLLCLLLGNLEPEAKLRVPLQRIHLGHRILNPLNGW
EQALSRSTSGSNPSPVPLSKFNTQFGTIGLTPQNFVYFDGSSNLWVPST
RCHFLSAPFWHFRNPKASSFRNGTKFAIQYGTGRSLSDNLTIGSIHDAFV
TFEALMPSLFFALAHFDGILGLGFLTAGVGPPLDAMVEOQLLEKPVFYLNR
DESGDGLSLVGGSDPAHYVPLPIFPIVTPAYQVHMSKVYGTGLSLCAQCSAI
LDGTSLITGPSEERALKNAIGGYPFLNGOYFIQCSPTPLPVPSHLGWNRLTG
ODYVTKILQSDVGLCLLGFQALDIPNAAGPLWLDVFLGYPVAVFDRGNVGRVG
LARAOSRSTDRRAETQAQFFKRRP"
BASE COUNT      263 a 384 c 356 g 328 t
ORIGIN

Query Match      38.3%; Score 789; DB 192; Length 1331;
Best Local Similarity 76.0%; Pred. No. 1.5e-207;
Matches 118; Conservative 0; Mismatches 305; Indels 16; Gaps 3;

QY 694 ajaaacggtcccccagatgtctccaccacgcgtctgcaacccctgctgctgctgtgc 753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 A|CTTAGTCCACAGATGTCCCACTACTGCTG-----CTGCTGCTGTCCTGC 52

QY 754 c|ctgctgaatgtgagccttcggggccacactgacatccgcatccctcttcatagatcc 813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53 T|CTGGGGAATTTGAGCCTGAGGAGGCCAACTGATCGTGTCCCTCTTCAACGAATCC 112

QY 814 a|u|ctggacgagggacccctgaacctactgagggatgagagaacacagacagctcccca 873
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 A|CTTGGACACAGATCTTAAACCCACTGAATGATGGAGACAGTGCAGAGACTTCTTA 172

QY 874 a|tttggggccccatccccctgggggcaagccccattctgtaacctctctcgaaactacaggg 933
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 -|-----GGACCTCCACCTCTGGTGCAACCCCTCTCTTGTGCTCTCTCCAACTCATGA 226

QY 934 a|gtgcagtattttgggaaattggctggaaacgctcccaaaacttactgtgct 993
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 227 A|ACCCAGTATTTGGAAGTATTTGGTTGGGAACCCCTCCCTCAGAAATTTACCCGTGTCT 286

QY 994 t|gaacatgctctcccaatctctgggtcccgctccagagagatgccactcttcagtgtgc 1053
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 287 T|GACACGGGTCTTCCAACTTGTGGTTCGGTCCACAGATGTCAATTTCTAGTTTG 346

QY 1054 c|ctgctgttacaccacgatttaaccgaagcctctagctctccaggcccaatggga 1113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 347 C|ATGCTGTTTCACCATCGCTTTAATCCAAAGCCTCCAGCTCTTCAGGCCCAATGGGA 406

QY 1114 c|aaagttgcatcaatataatggaactggcggttagatggaaatcctgagcgagacaagc 1173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 407 C|AAAGTTTGCCATTCAAGTATGGGACCGCGGCTGAGCGGAATCTGAGCCAGACAAATC 466

QY 1174 t|gaattgtgggaatcaagggtgcatcagtgatttcggggagagctctctggagacca 1233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 467 T|GATATCGGGGGATCCACGATCTTTGTGACATTTGGAGAGGCTCTGTGGAGGCCA 526

QY 1234 g|c|t|g|t|c|t|c|t|t|t|g|c|c|a|t|t|t|g|a|t|t|g|g|c|c|t|t|c|c|a|t|t|g|t 1293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 527 G|CCTGATCTTGTCTTAGCCACTTTGATGGGATCTTGGGCCCTGCGGCTTCCCACTCTGG 586

QY 1294 c|t|g|g|a|a|g|a|g|t|c|g|c|c|c|c|c|g|a|t|g|a|t|c|t|g|g|g|a|c|g|g|c|t|a|t|t|g|a|a|g|c 1353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 C|T|G|G|G|C|G|G|A|T|T|C|A|G|C|C|T|C|C|G|T|G|G|T|G|G|G|A|G|G|C|T|G|C|T|G|G|A|G|A|A|C 646

QY 1354 c|t|t|c|t|c|t|c|t|t|c|c|a|c|a|g|g|a|c|c|t|a|a|g|a|c|c|t|g|a|g|a|g|a|c|t|g|t|c 1413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 647 C|C|T|C|T|C|T|C|T|T|A|C|T|C|A|C|G|G|A|T|T|C|A|A|G|G|T|C|T|G|A|T|G|G|G|G|G|A|G|C|T|G|T|C 706

QY 1414 t|g|g|g|g|g|c|t|g|g|c|c|g|g|c|a|c|a|t|c|c|a|c|c|c|c|c|c|t|c|g|l|c|c|a|g|t|c|a|c|g|t|c 1473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 T|A|G|G|G|C|T|C|A|G|A|C|C|C|G|C|T|C|A|C|T|A|C|T|C|C|C|T|C|A|C|C|A|G|T|C|A|C|C|A|T|C 766

QY 1474 c|t|j|c|c|t|a|c|t|g|g|e|a|g|a|t|c|c|a|c|a|t|g|g|a|c|g|t|g|a|a|g|t|g|g|c|c|a|g|g|g|c|g|a|c|t|c|t|g|t|g 1533
```

```
Db 767 CTGCTACTGGCAGGTCCACATGGAGAGTGTGAAGTGGCAGAGGCTGAGCCTCTGTG 826
QY 1534 ccaaggctgtgtccactcctcctgatacggcagctcctcctcatcacagaccactaggg 1593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 827 CCACGGCTGCAGTCCCATCTTAGACACAGCAGCATCCTCATCACAGACTTAGTAGG 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1594 agatccggccctgcatcagccattgggggaatccctctgctggctggggagtagatca 1653
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 887 AGATCCGGCCCTTGAATAAAGCCATTTGGGGATATCCCTTCTGAATGGCAGTACTTCA 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1654 tcctgtgctcggaaatccaaagctccccgcagctcctctctctcttcttgggggggtcgtg 1713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 947 TTCAGTGTCCAAAGACGCCAAGCTTCCCTCTGCTCTCCACCTTGGTGGAGTCTGGT 1006
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1714 ttaacctcaagcccatgattacgtcatccagactactcgaatggcgcgcctctgct 1773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1007 TTAACCTCACAGGCAGGACTATGTATCAAGATTTCTACAGCGATCTTGGCCTCTGCC 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1774 tgcctgttccagccctggatgctcctcgcctcgcctcgcctcgcctcgcctcgcctcgc 1833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1067 TGTGGGCTTCCAAAGCCTTGGATATCCCAACGCTGCGGGACCCCTCTGATCCTTGGGG 1126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1834 acgtctcttgggacgtatgtggcgtcttcgacccgagggacatgaagacagcagccc 1893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1127 ACCTCTTTTGGGCCCTATGTGGCTGTCTTGACCGTGGGACAGAAAGGACTACGC 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1894 gggcgccctggcgcgctgcgcctgcgcctgcgcctgcgcctgcgcctgcgcctgcgc 1953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1187 GCGTGGGACTGGCGCTGCTCAGTCTCGTTCAACAGACCGCGCAGAAAGGACTACGC 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1954 agcgcgacttccccgggtgagcggcccaagtaa-agcgatcgcgagcggtggcgagag 2012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1247 AGGCGAGTCTTCAAAAGACGCGCTGTTAGGTACAGCTCACCGGCCACAGCAGCT 1306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2013 gtctgtaccaccagtaaaa 2031
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1307 ATGCTTCTTCCCAATATAA 1325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 2
BG540257 878 bp mRNA EST 03-APR-2001
DEFINITION 602569148F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4693859 5',
    mRNA sequence.
ACCESSION BG540257
VERSION BG540257.1 GI:13532490
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 878)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-r@mail.nih.gov
    Tissue Procurement: CLONTECH Laboratories, Inc.
    cDNA Library Preparation: CLONTECH Laboratories, Inc.
    DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
    Clone Distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    http://image.llnl.gov
    Plate: LCM1516 row: o column: 12
    High quality sequence stop: 718.
    Location/Qualifiers
        1..878
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4693859"
        /clone_lib="NIH_MGC_77"
```


High quality sequence stop: 452.
 Location/Qualifiers
 1. .612
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:244915"
 /clone_lib="NCL CGAP-Lu19"
 /tissue_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: p773D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 p773 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."

BASE COUNT 128 a 187 c 197 g 100 t
 ORIGIN

Query Match 29.4%; Score 605.2; DB 103; Length 612;
 Best Local Similarity 99.5%; Pred. No. 1.1e-156;
 Matches 507; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1440 a cccaccctcactctgtgcagtcacggtccctgctactgtgcagatccacatgag 1499
 |||||
 Db 612 A cccaccctcactctgtgcagtcacggtccctgctactgtgcagatccacatgag 553
 |||||
 QY 1500 cgtgtaagtgaggccaggtgactctctgtgcagaggtgtgtgcactctgtgat 1559
 |||||
 Db 552 CgtgtCAAGGTGGCCAGGGGTGACTCTGTGTGCAAGGGGTGTGCTGCTCTGGAT 493
 |||||
 QY 1560 a gggcagctccctcactacacagagaccactgagagatccggcctcatgcagcatt 1619
 |||||
 Db 492 A cggcagctccctcactacacagagaccactgagagatccggcctcatgcagcatt 433
 |||||
 QY 1620 g cgggaatccctctggtgggaggtacatcactgtgtcggaatcccaagctc 1679
 |||||
 Db 432 G cgggaatccctctggtgggaggtacatcactgtgtcggaatcccaagctc 373
 |||||
 QY 1680 c cgcagctctcctctctgtgtgggggtctgtgttaacctcacggcccatgattagtc 1739
 |||||
 Db 372 C cgtgcagctctcctctctgtgtgggggtctgtgttaacctcacggcccatgattagtc 313
 |||||
 QY 1740 atccagactactcgaatggcgcgcctctgtgtccggtttccagggccttgatgtc 1799
 |||||
 Db 312 A tccagactactcgaatggcgcgcctctgtgtccggtttccagggccttgatgtc 253
 |||||
 QY 1800 cctccgctcagggccctctgtatctcgtgaactcttcttgggagcgtatgtggcc 1859
 |||||
 Db 252 C cctccgctcagggccctctgtatctcgtgaactcttcttgggagcgtatgtggcc 193
 |||||
 QY 1860 gtcttcagcccggggacatgaagcagcgcgcgggtgggcctggcgcgctcgcact 1919
 |||||
 Db 192 G tcttcagcccggggacatgaagcagcgcgcgggtgggcctggcgcgctcgcact 133
 |||||
 QY 1920 c gtagcagacctcggatgggagagactcgcagcgcagttccccgggtgacgcca 1979
 |||||
 Db 132 C gtagcagacctcggatgggagagactcgcagcgcagttccccgggtgacgcca 73
 |||||
 QY 1980 ag gaaagcagcgcagcgggtgtgtcgcggaggtcctgctacccagtaaaatccacta 2039
 |||||
 Db 72 A gaaagcagcgcagcgggtgtgtcgcggaggtcctgctacccagtaaaatccacta 13
 |||||
 QY 2040 t tccattga 2049
 |||||
 Db 12 T t tccattga 3

RESULT 10

AL545162

LOCUS

DEFINITION

AL545162 LTI_NFL006.PL2 Homo sapiens cDNA clone CSOD1028YB18 5

prime, mRNA sequence.

ACCESSION

AL545162

VERSION

AL545162.1

KEYWORDS

GI:12877643

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 933)

AUTHORS

Li.W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. .933

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CSOD1028YB18"

/clone_lib="LTI_NFL006.PL2"

/tissue_type="placenta"

/note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 278 a 191 c 219 g 244 t

ORIGIN

1 others

Query Match 28.5%; Score 587.4; DB 106; Length 933;

Best Local Similarity 99.8%; Pred. No. 1.1e-151;

Matches 588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 117 tgaccagaaggttcaaaaacttcagcagcagctctgaagaatcgtgaaatgtgcg 176

|||

Db 1 TGACCAGAGAGGTTCAAAAACCTTCAGCGAGCTCTGAAGAACTGTGAAATGTGCGG 60

|||

QY 177 ttctcttaataaccagacaggttgaaaaattacagcaatattatatttttccctcta 236

|||

Db 61 TTTCTCTAAATACACAGACAGTTGAAAAATTACAGCAATATTATATTTTATTCCTCTA 120

|||

QY 237 aattcaagatcactcctgggttatctataaataaataaataaataaataaataaataa 296

|||

Db 121 AATTCAAGGATACCTACCTGGTGTATATTCTAAATGAATGGCTGAAACTCCTTTATGA 180

|||

QY 297 tatctgcagcactcctgaataataaccagacagcagcttctgactgcgaatacttggct 356

|||

Db 181 TATTCTGCAGCACCCTGTAATATACCAGAGAACAGCTTTGTCTACTGCGAAATCTGGCT 240

|||

QY 357 tcactgcacccccctccatggacaaatagtcagagtaagcgccctagatcccttaata 416

|||

Db 241 TCACGTCCATCCCCCTCCATGACAAATGAGTCAGAGTAAGCGCTAGGATCCCTTAATA 300

|||

QY 417 agtttaagcccaagcccggttcctctctctagcaactgacgttgcagcagcaggtttgg 476

|||

Db 301 AGTTTAAGCCCAAGCCCGCTCCATCTCTAGCAACTGACGTGGCCAGCCGAGGTTGG 360

|||

QY 477 acatacctcatgtagatgt 536

|||

Db 361 ACATACCTCATGTAGATGTGGTTGTCAACTTTGACATTCCTTACCATTCCTCAAGGATTA 420

|||

```
QY 537 tccatcgatgtagtgcacagctagagctggcgctcccgaaagcgtattactttgtca 596
|||||
Db 421 TCCATCGAGTAGTGCAGACAGTAGAGCTGGCGCTCCGGAAAGGCTATTACTTTGTCA 480

QY 597 cacagtatgatggaactcttccacgagcgcatagaacacttaattgggaagaactaccag 656
|||||
Db 481 CACAGTATGATGTGGAACCTCTCCAGCGCATAGAACACTTAATTTGGGAAGAACTACCCAG 540

QY 657 gttttcccaacagatgatgaggttatgatctgacaaacgcgtccc 705
|||||
Db 541 GTTTTCCCAACAGGATGATGAGGTATGATGCTGACAAACGCGTCCG 589

RESULT 11
LOCUS AI763426 762 bp mRNA EST 20-DEC-1999
DEFINITION wh92c02.x1 NCI_CGAP CLL1 Homo sapiens cDNA clone IMAGE:2388194 3'
similar to TR:009043 009043 KIDNEY-DERIVED ASPARTIC PROTEASE-LIKE
PROTEIN ; mRNA sequence.
ACCESSION AI763426
VERSION AI763426.1 GI:5179093
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 762)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 1254 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 476.
FEATURES
Location/Qualifiers
1..762
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2388194"
/clone_lib="NCI_CGAP CLL1"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCCATGCTTTTGTGTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 164 a 230 c 231 g 134 t 3 others
ORIGIN

Query Match 27.8%; Score 572.4; DB 24; Length 762;
Best Local Similarity 88.2%; Pred. No. 1.5e-147;
Matches 678; Conservative 0; Mismatches 84; Indels 7; Gaps 5;

QY 1281 tttccattctgtggaagagtgctggcccgatgtagtactgtgagcagggg 1340
|||||
Db 762 TTTCCATTCTGTGTGGAAGAGTGNCGGCCCGCTGGATGNACT-GTGAGACGAGG 704
```

```
QY 1341 ctattgataagcctgtctctctctcttccctcaacaggggaacctgaagcctgatgga 1400
|||||
Db 703 CTATGGGATTAAGCTGTCTCTCTTTTACTTCACANGGACCTTAAGTGGCTGATGGA 644

QY 1401 ggaagctgtctctgggggctggagcccggaacacatatacccccctcaccttcgtg 1460
|||||
Db 643 GAAGAGCTGGT-CTGGGGGGCTCAGACCC-GCACACTATACCCACCCCTCACCTTCGCTG 586

QY 1461 ccagtcacggtccctcctactactggaagatccacatgagcgtgtgaaggtggcccaagg 1520
|||||
Db 585 CCAGTCACAGTCCCGCCCTACT-GCAGATCCACATGGAGCGTGTGAAGTGGGCTCACGG 527

QY 1521 ctgaactctgtgccaagggtgtgtgcatctctgatacggcgacgtccctccatcaca 1580
|||||
Db 526 CTGACTCTCTGTGCCAGGGCTGTGTGTCATCTCGATACAGGCACACCTGTATCGTA 467

QY 1581 ggaacctgagagatccgggcccctgcatgcagccatgggggaatcccccttgcgtgct 1640
|||||
Db 466 GGACCCACTGAGGAGATCCGGGCCCTGCATGTCAGCCATTGGGGGAATCCCCCTTGCCTG 407

QY 1641 ggggaatatacatcctgtgctcggaatcccaaaagctcccccgagctctctctctt 1700
|||||
Db 406 TGGGAGTATCATATCCGGTGTCTCAGAAATCCCAAGCTCCCTCAGTCTCACCTCAT 347

QY 1701 ggggggtctgttttaacctcacggcccatgattacgtatccatccagactactcgaaatggc 1760
|||||
Db 346 GGGGGGCTGTGGTTAATCTCAGCGCCAGGATTACGTATCCAGTTGCTCAGGGTGAC 287

QY 1761 gtccgctctgtgttccggtttccagggcctggatgtccctccgctcgcagggcccttc 1820
|||||
Db 286 GTCCGCTCTGCTGTGTCGGCTTCCGGGCTTGAGACATCGTTCGGCTCCAGTACCTGTG 227

QY 1821 tggatcctggtagactcttcttggggagcgtatgtggcgtcttcgacccgaggacatg 1880
|||||
Db 226 TGGATCCTCGGCACGCTTTCTTTGGGGGCGTATGTACCGCTCTTCACCCGCGGGACATG 167

QY 1881 aagacagcgcccggtggcgctggcgctgcactcgagcgagcgacctcgatgg 1940
|||||
Db 166 AAGAGCGGCGCAGAGTGGGACTGGCGCGGCTCGCCCTCCGGAGCGGACCTGGGAAGG 107

QY 1941 ggagagactgcgagcgagtcctcccggtgagcggcccaagtgaagcgatgcgcagcg 2000
|||||
Db 106 CGGAGACCGCGCAGCGCAGTAGTACCGCGGTGCGGCCAGGTGATGCGATGCGATGCGCACCG- 46

QY 2001 gtggtcgagagctcctgctacccataaaatccactattccattga 2049
|||||
Db 47 --GGTAGCAGAGCTAGCGCTACTCAGTAAATAATCCAAATATTTCCATTGA 1

RESULT 12
LOCUS BF869544 587 bp mRNA EST 17-JAN-2001
DEFINITION IL3-ET0115-091000-286-E09 ET0115 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF869544
VERSION BF869544.1 GI:12259674
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 587)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.G.
```

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&t3=IL3-ET0115-091000-286-E03&t3-2000-10-09&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 563.

FEATURES

```

1. 38)
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cbone_lib="ET0115"
/dev_stage="Adult"
/note="Organ: Lung tumor; Vector: puc18; Site:1: Sma1;
Site:2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
164 g 161 g 145 t

```

BASE COUNT
ORIGIN

Query Match	27.3%	Score 562.2	DB 170	Length 587	0
Best Local Similarity	97.8%	Pred. No. 9.3e-145			
Matches 570	Conservative	0	Mismatches 13	Indels	Gaps
QY 724	cctgtctgaaccctctgctgtctgtcctctgtctgaatgtgagacctccggggcca	783			
Db	5 CCGTACCCCGCAACTGCTTGGATGCTTGCCTCTGCTGAATGTGGAGCTTCCGGGGCCA	64			
QY 784	cctgtatccgcatccctcttcatcgttcgaactcgaacgtggacgagaccctgaacctactga	843			
Db	65 CACTGATCCGATCCCTCTTCACTGAGTCCAACTGGACGCGAGGATCCTGAACCTACTGA	124			
QY 844	ggatggagagaaaccagcagagctccccaagttggggggcccatccoctggggacaagc	903			
Db	125 GCGATGGAGAGAACGACAGAGCTCCCAAGTTGGGGGCCCAATCCCTCGGGACAAGC	184			
QY 904	ccattctgtacctctctcgaactacagggatgtgcagtatTTTGGGAAATTTGGGCTGG	963			
Db	185 CCATCTTCGTACCTCTCTCGAACTACAGGGATGTGCAGTATTTTGGGAAATTTGGGCTGG	244			
QY 964	gaacgctccacaaaacttcaactgttcctttgacactggctctccaatctctgggtcc	1023			
Db	245 GACGGCTCCACAAAATTTCACTGTTGCCCTTGGACCTGGCTCCTCCAATCTCTGGGTCC	304			
QY 1024	cgctcagagatgcgaacttcttcagttgacctgtggttacaccacccgattgatccca	1083			
Db	305 CGTCCAGGAGATGCCATTCTTCACTGTGCCCTGCTGGTTACACCCGATTTGATCCCA	364			
QY 1084	aagcctctagctcttcacggccaatgggaccaagttggcattcaatatggaactgggc	1143			
Db	365 AA5CCTCTAGCTCCTTCCAGGCCAATGGGACCAAGTTTGGCATTCATAAT7GGAACCTGGGC	424			
QY 1144	ggtagatggaatcctgagcagagcaagctgactattgagaaatcaaggggtgcacag	1203			
Db	425 GG3TAGATGGAATTCCTGAGCGAGGACAAGCTGACTATTGGTGGAAATCAAGGGTGATCAG	484			
QY 1204	tgaatttcggggagctctctggagccagacctggtcttctcgcttttgccatttgatg	1263			
Db	485 TGATTTTCGGGGAGGCTCTCTGGAGCCACGCTTGGTCTTGGCTTTTGCCCATTTTGATG	544			
QY 1264	ggaatttgggctcggtttttcccatctctctgtggaagaggt	1306			

Db 545 GGATATTGGGCTCGGTTTCCAAATCTGTCTGTGGAAGGAGT 587

RESULT 13
BE892689

BE892689. 880 bp mRNA EST 20-OCT-2000
 60143307F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3919273 5',
 mRNA sequence.
 BE892689
 BE892689.1 GI:10353284
 EST.
 human.

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 880)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
Contact: Robert Strausberg ph.D.

COMMENT
Contact: Robert Strausberg, Ph.D.
Unpublished (1999)

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ARCC/DCMP/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM9748 row: m column: 02
High quality sequence stop: 671.

FEATURES	
SOURCE	

```

1. ;000
/organism="Homo sapiens"
/db xref="taxon:9606"

```

/db_xref="taxon:9606"
/clone="IMAGE:391973"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic mel"
/lab_host="PH10B (phage-res)
/note="Organ: skin; Vector:
Site_2: Sali; Cloned unidir
Average insert size 2 kb.
Technologies."

BASE COUNT	257 a	197 c	205 g	221 t
ORIGIN				

Query Match	Score	DB	Length
Best Local Similarity	27.1%	558.2	141
Best N-gram Similarity	90.3%	558.2	141

Best local similarity 90.3%; pred. NO. 1.4e-143;
Matches 649; Conservative 0; Mismatches 48; Indels 22; Gaps 4

QY 1 cttgagagctctcaataacttggtcatggatgaagcgaccgaatactgaatatggattt 60

Db
51 CTTGAGAGCTCTCAATACTTGGTCATGGATGAGCGGACCGAATACTGAATATGGATT 110

QY 61 tgagacagaggttgacaag-----cctcgagatcggaacattctctt 105

Db
111 TGACACAGAGGTGACACAGATCCTCAAAGTGATTCTTCGAGATCGGAAACATTCTCTTT 170

QY 106 ctctgcaccatgaccaagaagggttcaaaaaaciltcaggcgagcagctctgaaagaatcctgt 165

Db 171 CTCTGCCACCATGACCAAGAAGGTTCAAAAATTTCAGCGGACGCTCTGAAGAAATCCTGT 230

Qy 166 gaaatgtgcggtttcctctaaataccagacadtggaaaaattacagcaatatatatattt 225

Db

Qy 226 tattccctctaaattcaaggatacctacctggtttatatattctaaatgaattgcctgaaaa 285

291 TATTTCCCTCTAAATTCAAGGATACCTACCTGGGTTTATATTTCTAAATGAATTGGCTGGAAA
Db

QY 286 ctccctttatgatattcttcgcagcaactgtataataaccagagagaacacgttttgcactacgc 3455

Db
351 CTCTTTTATGATATCTGCAGGACCTGTAAATAATACCCAGACAACAGCTTTTCTACGCC
|||||
352
353
354

FEATURES

source: 1..837
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1284501"
/clone_lib="NCI CGAP GCBI"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Alliman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTCACCAATGAAGTGGAGCGCGCCCTCATTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
185 a 236 c 255 g 160 t 1 others

BASE COUNT
ORIGIN

Query Match 26.7%; Score 550; DB 11; Length 837;
Best Local Similarity 84.7%; Pred. No. 2.5e-141;
Matches 712; Conservative 0; Mismatches 116; Indels 13; Gaps 8;

QY 1222 ttgggagccagctggttctgttccatttggatggatattgg-gcctcggt 1280
Db 830 TTTGGAAAACCAACCTTGCTTCACTGTTCCTCCGCCCTGGGATATTGGCGCCTCGTT 771

QY 1281 ttcccatctctgtgaaagagttcggcccccagatgagtactggtggagcagggg 1340
Db 770 TTTCCCATTTCTGTGTGGAAGATTTTCGCCCCCTGGATTACTGGTGGAGCAGGGC 711

QY 1341 ctattgataagcctgtcttctcttctacotcaacagagaccctgaagaccctgagatgga 1400
Db 710 CTATTGGATAAGCCTGTCTCTCCCTTACTTCAACAGGCACCCCTGAAGTGGCTGAT-GA 652

QY 1401 gtagagctggtcctgggggctcggaccgacactacatccca-cacctcaactcgt 1459
Db 651 GATAGAGCTGGTCTCTGGGGCCCTCAGACCCCGCACACTACCCACCCCTCAGCTTCGT 592

QY 1460 gacagtcaggtccctgctactgcagatccacatgagcgtgtgaggtgggcccagg 1519
Db 591 G(CAGTCACAGTCCCGCCTACTTGGCAGATCCACATGAGCGTGTGAGGTGGGCTCAGC 532

QY 1520 gctgactctctgccaaggcgtgctgctcctcctgatacgggacgctccctcatcac 1579
Db 531 GGTGACTCTCTGTGCCAGGGCTGTGTGCTCATCTTGGATACAGGCACACCTGTCTCGT 472

QY 1580 aggacccactgagagatccgggcccctgcatgagccattgggggaatcccccttgcggc 1639
Db 471 AGGACCCACTGAGGAGATCCGGGCCCTGCATGACAGCCATT-GGGGAATCCCTTGTGTCG 413

QY 1640 tgggagtaacatacctctgtctcggaaatcccaagctcccccagctctcctctctct 1699
Db 412 TGGGGAGTACATATCCGGTGTCTCAGAAATCCCAAGCTCCCGCAGTCTCACTCCCTCAT 353

QY 1700 tgggggggtctggttaacctcagcccatgattacgtcatccagactactcgaaatgg 1759
Db 352 TGCGGGGTCTGGTTTAATCTACGGGCCAGGATTAAGTCATCCAGTTTGTCTCAGGGTGA 293

QY 1760 cgtccgcctctgtctccggtttccaggccctggatgtctccctccgctcagggccctt 1819
Db 292 CGTCCGCTCTGTCTGTCCGGCTTCGGGGCTTTGGA-CATCTTCGCTCCACTACTGT 234

QY 1820 ctggactcgtgagcgtctcttgggaagcgtatgtggcgtcttcgacccggggacat 1879
Db 233 GTTAATCTCGGCAGCGTTTCTTGGGGGGCGTATGTGACCGTCTTCGACCGGGGACAT 174

QY 1880 gattgagcagcgcgcgggtgggctggcgctgcactcagcagcagcagcactcggatg 1939

Search completed: November 17, 2001, 13:34:21
Job time: 3420 sec

Db 173 GAAGAGCGGCGCACGAGTGGGACTGGCGCGCTCGCCCTCCGGAGCGGACCTGGGAAGC 114
QY 1940 gggagagactgcgcagggcagttcccccgggtgacgcccagaagtgaagcgcagtcgcagcg 1999
Db 113 GCGAGACA---CGCAGGCGCAG-TACCCCGGTGCCCGCCAGGTTGNATGTCATGCGCACCG 58

QY 2000 ggtggtcgcggagggtcctgctaccacagtaaaaaatccactatttccattgaaaaaataaaa 2059
Db 57 GGTAGCAGAGTAG----CGCTACTCAGTAAAAAATCCAATATTTCCATTTGAAAAAATAAAA 2
QY 2060 a 2060
Db 1 A 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:54:41 ; Search time 4216.79 Seconds
(without alignments)
1991.797 Million cell updates/sec

Title: US-09-700-770-6
Perfect score: 543
Sequence: 1 ccggcgctgagggcgagg.....gttaagagcaaaaaaaaaa 543

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: em_ba1.*
17: em_ba2.*
18: em_fun.*
19: em_htgo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_other.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vl.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vl1.*
59: gb_vl2.*
60: gb_vtg1.*
61: gb_vtg2.*
62: gb_vtg3.*
63: gb_vtg4.*
64: gb_vtg5.*
65: gb_vtg6.*
66: gb_vtg7.*
67: gb_vtg8.*
68: gb_vtg9.*
69: gb_vtg10.*
70: gb_vtg11.*
71: gb_vtg12.*
72: gb_vtg13.*
73: gb_vtg14.*
74: gb_vtg15.*
75: gb_vtg16.*
76: gb_vtg17.*
77: gb_vtg18.*
78: gb_vtg19.*
79: gb_vtg20.*
80: gb_vtg21.*
81: gb_vtg22.*
82: gb_vtg23.*
83: gb_vtg24.*
84: gb_vtg25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_ro1.*
95: gb_ro2.*
96: gb_in4.*
97: gb_pr10.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	471	86.7	471 97 HUM2B52D10
C 2	241.4	44.5	96482 67 AC022095
C 3	241.4	44.5	168347 69 AC025336
C 4	60.4	11.1	40104 3 SCF11
C 5	60.2	11.1	34611 3 SCD63A
C 6	56	10.3	208936 61 AC010821
C 7	55.8	10.3	166539 62 AC011402
C 8	55.4	10.2	33001 3 SCC123
			AF086152 Homo sapi
			AC022095 Homo sapi
			AC025336 Homo sapi
			AL132662 Streptomy
			AL356832 Streptomy
			AC010821 Homo sapi
			AC011402 Homo sapi
			AL136518 Streptomy

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6686

Center clone name: 451_H_23

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 150422 bases at least Q40

Consensus quality: 159524 bases at least Q30

Consensus quality: 163013 bases at least Q20

Insert size: 165247; sum-of-contigs

Quality coverage: 3.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1389: contig of 1389 bp in length
1390 1489: gap of 100 bp
1490 3130: contig of 1641 bp in length
3131 3230: gap of 100 bp
3231 4942: contig of 1712 bp in length
4943 5042: gap of 100 bp
5043 6981: contig of 1939 bp in length
6982 7081: gap of 100 bp
7082 8708: contig of 1627 bp in length
8709 8808: gap of 100 bp
8809 10286: contig of 1478 bp in length
10287 10386: gap of 100 bp
10387 12212: contig of 1826 bp in length
12213 12312: gap of 100 bp
12313 14658: contig of 2346 bp in length
14659 14758: gap of 100 bp
14759 17941: contig of 3183 bp in length
17942 18041: gap of 100 bp
18042 21297: contig of 3256 bp in length
21298 21397: gap of 100 bp
21398 24992: contig of 3595 bp in length
24993 25092: gap of 100 bp
25093 27768: contig of 2676 bp in length
27769 27868: gap of 100 bp
27869 31188: contig of 3320 bp in length
31189 31288: gap of 100 bp
31289 33714: contig of 2426 bp in length
33715 33814: gap of 100 bp
33815 37277: contig of 3463 bp in length
37278 37377: gap of 100 bp
37378 42302: contig of 4925 bp in length
42303 42402: gap of 100 bp
42403 47816: contig of 5414 bp in length
47817 47916: gap of 100 bp
47917 52586: contig of 4670 bp in length
52587 52686: gap of 100 bp
52687 56567: contig of 3881 bp in length
56568 56667: gap of 100 bp
56668 61557: contig of 4890 bp in length
61558 61657: gap of 100 bp
61658 66724: contig of 5067 bp in length
66725 66824: gap of 100 bp
66825 71568: contig of 4744 bp in length
71569 71668: gap of 100 bp
71669 76578: contig of 4910 bp in length
76579 76678: gap of 100 bp
76679 83312: contig of 6634 bp in length
83313 83412: gap of 100 bp

```

```

* 83413 90053: contig of 6641 bp in length
* 90054 90153: gap of 100 bp
* 90154 99426: contig of 9273 bp in length
* 99427 99526: gap of 100 bp
* 99527 108015: contig of 8489 bp in length
* 108016 108115: gap of 100 bp
* 108116 118144: contig of 10029 bp in length
* 118145 118244: gap of 100 bp
* 118245 130468: contig of 12224 bp in length
* 130469 130568: gap of 100 bp
* 130569 142239: contig of 11671 bp in length
* 142240 142339: gap of 100 bp
* 142340 157135: contig of 14795 bp in length
* 157136 157235: gap of 100 bp
* 157236 168347: contig of 11112 bp in length.
FEATURES
      source
      1..168347
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="5"
      /map="5"
      /clone="RP11-451H23"
      /clone_lib="RPC1-11 Human Male BAC"
      1..1389
      /note="assembly_fragment"
      1490..3130
      /note="assembly_fragment"
      3231..4942
      /note="assembly_fragment"
      5043..6981
      /note="assembly_fragment"
      7082..8708
      /note="assembly_fragment"
      8809..10286
      /note="assembly_fragment"
      10387..12212
      /note="assembly_fragment"
      12313..14658
      /note="assembly_fragment"
      14759..17941
      /note="assembly_fragment"
      18042..21297
      /note="assembly_fragment"
      21398..24992
      /note="assembly_fragment"
      25093..27768
      /note="assembly_fragment"
      27869..31188
      /note="assembly_fragment"
      31289..33714
      /note="assembly_fragment"
      33815..37277
      /note="assembly_fragment"
      37378..42302
      /note="assembly_fragment"
      42403..47816
      /note="assembly_fragment"
      47917..52586
      /note="assembly_fragment"
      52687..56567
      /note="assembly_fragment"
      56668..61557
      /note="assembly_fragment"
      61658..66724
      /note="assembly_fragment"
      66825..71568
      /note="assembly_fragment"
      71669..76578
      /note="assembly_fragment"
      76679..83312
      /note="assembly_fragment"
      83413..90053
      /note="assembly_fragment"

```

```

misc_feature      90154..99426
                  /note="assembly_fragment"
misc_feature      99527..108015
                  /note="assembly_fragment"
misc_feature      108116..118144
                  /note="assembly_fragment"
misc_feature      118245..130468
                  /note="assembly_fragment"
misc_feature      130569..142239
                  /note="assembly_fragment"

Query Match      44.5%; Score 241.4; DB 69; Length 168347;
Best Local Similarity 99.6%; Pred. No. 5.3e-32;
Matches 242; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 142 ccgctgctgtttcttaagtgcctgcgcgaagcctgtgcccagcctgctgctgcgtgg 201
      |||
Db 144047 CAGCTGCTGCTTTCTTATGTGGCTGCGGCAAGCCCTGTGCCAGCCTGTGCTGGCGTGG 143988

Qy 202 agtcggcgagcgagccggggccgggaacccctggccaaacccctcgccaccctcaaccgcg 261
      |||
Db 143987 AGTGGCGGCGAGCGCGGCGGCGGACCCCTGGCCAAACCCCTCGGCACCCCTCAACCCGC 143928

Qy 262 tgaagctcctgctgagcagcctggtggaatccctcggaacacacotcagagggctcccgaga 321
      |||
Db 143927 TGAAGCTCCTGCTGAGCAGCCTGGGCATCCCGCTGAACCACTCATAGAGGGCTCCCGAGA 143868

Qy 322 agtgtgtgctgagctggtggtcccaagcgctggggccggtgaagccctgaagccctgcg 381
      |||
Db 143867 AGTGTGTGGCTGAGCTGGTGTCCCGAGCGCGTGGGGCGCGTGAAGGCCCTGAGGCCCTGC 143808

Qy 382 tgg 384
      |||
Db 143807 TGG 143805

RESULT 4
LOCUS      SCF11 40104 bp DNA BCT 26-OCT-1999
DEFINITION Streptomyces coelicolor cosmid F11.
ACCESSION AL132662
VERSION AL132662.1 GI:6137022
KEYWORDS   3-oxoacyl-(acyl-carrier-protein) synthase; acyl carrier protein;
acyltransferase; alpha-galactosidase; beta-hexosaminidase; cold
shock protein; gluconolactonase precursor; histidine kinase
protein; hydrolase; oxidoreductase; peptidase; pyruvate
carboxylase; regulatory protein; response regulator; sugar
hydrolase; sugar transporter; transcriptional regulator;
transmembrane protein.
SOURCE      Streptomyces coelicolor A3(2).
ORGANISM    Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces.
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 40104)
Redenbach,M., Kieser,H.M., Denapate,D., Eichner,A., Cullum,J.,
Kinashi,H., and Hopwood,D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
2 (bases 1 to 40104)
Seeger,K.J. and Harris,D.
Unpublished
3 (bases 1 to 40104)
Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
Direct Submission
Submitted (26-OCT-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded

```

by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid F11 overlaps with cosmid F6 on the AseI-F genomic restriction fragment.

FEATURES

Location/Qualifiers	source
1..40104	
/organism="Streptomyces coelicolor A3(2)"	
/strain="A3(2)"	
/db_xref="taxon:100226"	
/clone="cosmid F11"	
1..100	gene
/gene="SCF11.01"	
<1..100	CDS
/gene="SCF11.01"	
/note="SCF11.01, possible tetR family transcriptional regulator, partial CDS, len: >32 aa. Previously annotated as Streptomyces coelicolor SCF6.17, possible tetR family transcriptional regulator, len: >193 aa. Similar to several e.g. Streptomyces coelicolor TR-CAB56142 (EMBL; AL117669) putative transcriptional regulator SCF12.15 (188 aa), fasta scores opt:446 z-score: 521.2 E(): 1.2e-21 43.5% identity in 184 aa overlap and Streptomyces glaucescens SW:TCMR_STRGA (EMBL; M80674) tetracenomycin C transcriptional repressor (226 aa), fasta scores opt: 202 z-score: 242.5 E(): 3.8e-06 29.3% identity in 164 aa overlap."	
/codon_start=2	
/transl_table=11	
/product="putative tetR family transcriptional regulator (partial)"	
/protein_id="CAB59578.1"	
/db_xref="GI:6137023"	
/translation="IDGGERPLRLPLLAECFDVLTLSARGAASAG"	
1..100	misc_feature
/gene="SCF11.01"	
/note="nominal overlap with Streptomyces coelicolor cosmid StF6"	
254..257	RBS
276..1067	gene
/gene="SCF11.02"	
276..1067	CDS
/gene="SCF11.02"	
/note="SCF11.02, probable oxidoreductase, len: 262 aa; similar to TR:O54197 (EMBL:AJ000671) Streptomyces clavuligerus clavulanate-9-aldehyde reductase, 247 aa; fasta scores: opt: 501 z-score: 567.7 E(): 3e-24; 39.1%	

	identity in 238 aa overlap. Contains Pfam match to entry PF00106 adh_short, short chain dehydrogenase and prosite match to entry PS0061 Short-chain dehydrogenases/reductases family signature. also similar to Streptomyces coelicolor SCF6.15c; 245 aa; fasta scores: opt: 818 z-score: 805.2 E():0; 56.5% identity in 237 aa overlap" /codon_start=1 /transl_table=11 /product="putative oxidoreductase" /protein_id="CAB59579.1" /db_xref="GI:6137024" /translation="MSRFDSSQQRMTTHRKDVMQTQSKVVLVTGASSGIGEATLRLLAADGHRFLGARRTERLEKLAARIAEDGGTAGYRRLDTDAADVRAFSVAARWGLALDVIYVNNAGVSPLESLKLAARMDLVNRYGLVHGTAAALPVMRAQGGGHVVNTLGVAGVEPTAAVYCATFAVRALSEGRLQESAGDIRVSYSPGVTSSELADSIDSPRAEDKMTTYSVAVPAASATADATAFASVSPAEDVNEIVVRPAASAQ" 348..899 /gene="SCF11.02" /note="Pfam match to entry PF00106 adh_short, short chain dehydrogenase, score 248.20, E-value 1.2e-70" 753..839 /gene="SCF11.02" /note="PS00061 Short-chain dehydrogenases/reductases family signature" complement(1095..1532) /gene="SCF11.03c" complement(1095..1532) /gene="SCF11.03c" /note="SCF11.03c, probable transcriptional regulator, len: 145 aa; similar to SW:ASNC_ECOLI (EMBL:AE000451) Escherichia coli, regulatory protein AsnC, 152 aa; fasta scores: opt: 222 z-score: 287.1 E(): 1.3e-08; 26.9% identity in 145 aa overlap. Contains Pfam match to entry PF01037 ASNC_trans_reg, AsnC family" /codon_start=1 /transl_table=11 /product="putative transcriptional regulator" /protein_id="CAB59580.1" /db_xref="GI:6137025" /translation="MDEIDRRLLALQQDAPRYAALGGEVGLSAGAAHRYVRKLRREQKVRITTVDPBALDGRGVLAFLVVDSTAMMGDRAEDFAAPETIOFAHVIAGSAALVQKVTATRQLRDVLRRLYAIIDGVSQTATVLTETFERPLDAG" complement(1164..1460) /gene="SCF11.03c" /note="Pfam match to entry PF01037 ASNC_trans_reg, AsnC family, score 52.20, E-value 1.1e-11" 1608..1612 1617..2525 1617..2525 /gene="SCF11.04" /gene="SCF11.04" /note="SCF11.04, possible gluconolactonase, len: 302 aa; similar to SW:GNL_ZYMMO (EMBL:X67189) Zymomonas mobilis gluconolactonase precursor (EC 3.1.1.17) D-glucono-delta-lactone lactonohydrolase; fasta scores: opt: 326 z-score: 392.6 E(): 1.7e-14; 28.7% identity in 258 aa overlap" /codon_start=1 /transl_table=11 /product="possible gluconolactonase precursor" /protein_id="CAB59581.1" /db_xref="GI:6137026" /translation="MSTDGPYRILDDFRFTGRCANGNRLEVLVDGCRWAGPLVLPALWQLVMSDIPNDRILRDLRDEATGSGVFRAPAGHSNGNTLDRQGLVTCQGNRVVTRT EPDGRVTVLAERFDGRRLSNPNDVSVRSDPTGFWSDFDPGFTSYEGHRASEIGACN VYRIDPVSQGVRLADGDEGPGNGVILITPDERRLFVSDSRAAKIHAFDIREDTLSLDGK VRAEDGQVHFENIRFDEGRLWALHDGHVCHYDPTDGTLTGLRLRVPPEVSNVAFGG KNRRLFTATTSLYSLWVSZAPRL" 2656..3471 /gene="SCF11.05" 2656..3471 /gene="SCF11.05"
misc_feature	
misc_feature	
gene	
CDS	
misc_feature	
RBS	
gene	
CDS	
gene	
CDS	

REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 34611)
Brown, S.P. and Harris, D.
Unpublished
3 (bases 1 to 34611)
Cerdano, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
Direct Submission
Submitted (24-MAY-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK

NOTES:

Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.

(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)

CDS are numbered using the following system eg SC7B7.01c, SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).

The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.gov/jp/>

jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or (att)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.

FEATURES

source Location/Qualifiers
1..34611
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid D63A"
complement(1..106)
/gene="SCD63A.01c"
misc_feature 1..100
/note="nominal overlap qwith Streptomyces coelicolor
cosmid SCD63"
complement(<1..106)
/gene="SCD63A.01c"
/note="SCD63A.01c, possible integral membrane protein
(fragment), len: >35 aa"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB92653.1"
/db_xref="GI:8218191"
/translation="MTDHEHAATGPGGSGPRPPQGAGPQDAAPADAG"
297..1589
/gene="SCD63A.02"
297..1589
/gene="SCD63A.02"
/note="SCD63A.02, possible two-component system sensor
kinase, len: 430 aa; C-terminal domain similar to
TR:CA51969 (EMBL:AL109661) Streptomyces coelicolor
putative two-component system sensor kinase SC6E10.15c,
413 aa; fasta scores: opt: 168 z-score: 198.3 E(): 0.0014;

32.7% identity in 248 aa overlap. Contains possible
hydrophobic membrane regions in N-terminal domain"
/codon_start=1
/transl_table=11
/product="putative two-component system sensor kinase"
/protein_id="CAB92654.1"
/db_xref="GI:8218192"
/translation="MPEAAAPVVEPPRPKLYRSSDGRWLGGVARGLAGHLGLPVVW
VPFAFGVLPMDGLGALLYAAFWFVPLGVGGVEAKPPSLVTSFETPCDGRRLVTRK
KRGQIVALLLMVVALVFCVNDLGGAKAYLWPTVLVGAGVALVWRQDNRARRW
AEVGRHRTVTLRSVGGVLLVTAGVTGIFVLOGSAAHLGSVLOALAVLVGITLLAG
PYLVMTQDLSEERLMRIRAEAEAAHVHDSVLHTLTLLIQRNAENAGVRRRLAAQ
ERDLRSWLYKEPNQGDDEDDPTLAEAVKRNAAEVEKKGVLPEVVVVGDCPLDERI
SAQMQAAREAMNAAYKGEGGQVQVVAEVEGRTVFVSVRDRGPGFDFDSTPADRMGV
RESIIIGRMERNGGTARLRAVPDGGTVELEMERAEKTS"
1574..1578
/gene="SCD63A.02"
1586..2341
/gene="SCD63A.03"
1586..2341
/gene="SCD63A.03"
/note="SCD63A.03, probable two-component system
DNA-binding response regulator, len: 251 aa; similar to
SW:NARL_ECOLI (EMBL:X13360) Escherichia coli
nitrate/nitrite response regulator protein NarL, 216 aa;
fasta scores: opt: 428 z-score: 490.1 E(): 7.8e-20; 34.1%
identity in 214 aa overlap and to TR:CA51968
(EMBL:AL109661) Streptomyces coelicolor putative
two-component system regulator SC6E10.14c, 215 aa; fasta
scores: opt: 517 z-score: 472.3 E(): 9.4e-21; 44.3%
identity in 219 aa overlap. Contains Pfam matches to
entries PF0072 response_reg, Response regulator receiver
domain and PF00196 GerE, Bacterial regulatory proteins,
luxR family and match to Prosite entry PS00622 Bacterial
regulatory proteins, luxR family signature. Contains also
possible helix-turn-helix motif at residues 206..227
(+2.85 SD)"
/codon_start=1
/transl_table=11
/product="putative two-component system DNA-binding
response regulator"
/protein_id="CAB92655.1"
/db_xref="GI:8218193"
/translation="MSDPTANGTAGTAGAGAAAPQAQAGDGAGERRVRVLLVDDHR
MFTGVQAEIGQTAEVGVVGEAADVDQAVTVITARPEVLLDHLPGCGGVLELR
RCAPLMSDAERPVRFALSVDAAEDVIGVIRGARGVFTKTITGLDVSVERVQSG
DAVFSPLAGFVLDAFASTDAPPVDELDRLTOREREVLRIARGAYKEIAKQLYIS
VKTVESHSVAVLRKRLQLSNRHELTWRATARRLV"
1688..2047
/gene="SCD63A.03"
/note="Pfam match to entry PF0072 response_reg, Response
regulator receiver domain, score 55.50, E-value 1.2e-12"
2147..2338
/gene="SCD63A.03"
/note="Pfam match to entry PF00196 GerE, Bacterial
regulatory proteins, luxR family, score 73.60, E-value
4.2e-18"
2198..2281
/gene="SCD63A.03"
/note="PS00622 Bacterial regulatory proteins, luxR family
signature"
complement(2443..3486)
/gene="SCD63A.04c"
complement(2443..3486)
/gene="SCD63A.04c"
/note="SCD63A.04c, probable NPL/P60 family secreted
protein, len: 347 aa; similar to TR:Q9XAO3 (EMBL:AL078618)
Streptomyces coelicolor SCD16A.22, 277 aa; fasta scores:
opt: 409 z-score: 418.8 E(): 7.3e-16; 34.8% identity in
250 aa overlap, to TR:Q9X725 (EMBL:AL049497) Streptomyces
coelicolor putative secreted protein SC6G10.09c, 347 aa;
fasta scores: opt: 928 z-score: 800.9 E(): 0; 44.2%
identity in 353 aa overlap and to Streptomyces coelicolor

correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid C123 lies between and overlaps with cosmids C77 and C88 on the AseI-C genomic restriction fragment.

FEATURES

```

source
1..33001
  /organism="Streptomyces coelicolor A3(2)"
  /strain="A3(2)"
  /db_xref="taxon:100226"
  /clone="cosmid C123"

misc_feature
1..104
  /note="nominal overlap with cosmid stc77 between bases
3398..33801."
  /gene="rspT"
  /gene="rspT"
  /gene="rspT"
  /note="SCC123.01, rspT, 30s ribosomal protein S20, len: 88
aa. Highly similar to Bacillus subtilis
SW:RS20_BACSU(EMBL:D84432) 30S ribosomal protein S20
(BS20) (87 aa), fasta scores opt: 226 z-score: 298.8 E():
3e-09 44.0% identity in 84 aa overlap and Mycobacterium
tuberculosis SW:RS20_MYCTU(EMBL:Z81368) 30S ribosomal
protein S20 (86 aa), fasta scores opt: 363 z-score: 465.7
E(): 1.5e-18 70.9% identity in 86 aa overlap. Contains a
-pfam match to entry PF01649 Ribosomal_S20p, Ribosomal
protein S20."
  /codon_start=1
  /transl_table=11
  /product="30s ribosomal protein S20."
  /protein_id="CAB66241.1"
  /db_xref="GI:6714669"
  /translation="MANIKSQIKRNTNEKARLNKSKVSLKTAIRKAREAAAGDV
EKATFEQVARSRELDKAVSGKVIHKQNQKKSALAKVGALKG"
  /gene="rspT"
  /note="Pfam match to entry PF01649 Ribosomal_S20p,
Ribosomal protein S20, score 129.10, E-value 7.9e-35"
  /complement(583..1593)
  /gene="SCC123.02c"
  /complement(583..1593)
  /gene="SCC123.02c"

misc_feature
12..263
  /note="Pfam match to entry PF01649 Ribosomal_S20p,
Ribosomal protein S20, score 129.10, E-value 7.9e-35"
  /complement(583..1593)
  /gene="SCC123.02c"
  /complement(583..1593)
  /gene="SCC123.02c"

gene
  /note="SCC123.02c, possible DNA-binding protein, len: 336
aa. Highly similar to several proteins of undefined
function e.g. Mycobacterium tuberculosis
TR:P71730(EMBL:Z81368) hypothetical 33.1 kD protein (316
aa), fasta scores opt: 792 z-score: 880.3 E(): 0.42.4%
identity in 316 aa overlap. Note that codon 21 may be an
alternative translational start site. Contains a possible
helix-turn-helix motif situated between residues 263..284
(+2..96 SD)."
  /codon_start=1
  /transl_table=11
  /product="putative DNA-binding protein."
  /protein_id="CAB66242.1"
  /db_xref="GI:6714670"
  /translation="MRMLVAMARKTANDDPLAPVTLAVGQEDLLLDRAVOEVVAAAK
AADADYDRLTPOLOPCTLAELTSPSLFAERKVVVVRNAQDLSDATVYKDVKAYLGA
PAEETIVLLHAGAKGKGLDAGKAGKACVACPKTKPADRLAFVRAEFRTAGRSA
TPEACQALVDAIGSLRELASVQLTADVEGTDEAVVGYRTGRASFTVADRA
VEGRAAEALRWLSLATGCVAPLITSALAQGVRAIGKLSARGGRPADLARELGMP
WKIDRVQOMRGWTPDGVSVALRAVAEADAGVKGGDDPEVALEKAVVVIARAARSRG
RT"
  /complement(1638..1859)
  /gene="SCC123.03c"
  /complement(1638..1859)

gene
  /note="SCC123.03c, possible DNA-binding protein, len: 355
aa. Similar to Mycobacterium tuberculosis
TR:P71728(EMBL:Z81368) hypothetical 30.5 kD protein (297
aa), opt: 271 z-score: 260.6 E(): 4e-07 36.3% identity
in 292 aa overlap. Also similar in parts to Bacillus

```

```

  /gene="SCC123.03c"
  /note="SCC123.03c, unknown, len: 73 aa."
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein SCC123.03c."
  /protein_id="CAB66243.1"
  /db_xref="GI:6714671"
  /translation="MLPRDAAEQAAQELGERFGTDEEPRILRDALAGEDDAEDAQWL
VLLHDERALDPGELDAFAGEWEGWREP"
  /complement(1915..2796)
  /gene="SCC123.04c"
  /complement(1915..2796)
  /gene="SCC123.04c"
  /note="SCC123.04c, unknown, len: 293 aa. Similar in parts
to several proteins of unknown function e.g. Streptomyces
coelicolor TR:O86515(EMBL:AL031124) hypothetical 20.0 kD
protein (185 aa), fasta scores opt: 384 z-score: 456.7
E(): 4.7e-18 32.8% identity in 180 aa overlap."
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein SCC123.04c."
  /protein_id="CAB66244.1"
  /db_xref="GI:6714672"
  /translation="MIGRLGSRTRKMQRTGPLAAVHTPPDAGVLSRVLDVNPVPT
HAETVSTMGKVVGGTDPGSEFVATPAGCYRLAVSAEGYTPYRSVTVAEPTLA
SLGDVTLQVAPPPELPAPGDWEIEPAHSIAFTARHIGLARIHGRFNSFAGAVRTAD
MEQSAMHVVIDAASIDITNVKMRDHLRSADFLDQRYPTLEFYSDRFARHGRNAV
GALSILGVTRTVTLDTVEYLGNGMEGETRAACRATTELHRDDFTVSNQTMTRGI
VGPSIRIDLDVQTVPKG"
  /complement(2806..2809)
  /complement(3193..5760)
  /gene="SCC123.05c"
  /complement(3193..5760)
  /gene="SCC123.05c"
  /note="SCC123.05c, possible integral membrane protein,
len: 855 aa. Similar to Bacillus subtilis
SW:CME3_BACSU(EMBL:L15202) comE operon protein 3, required
for the binding and uptake of transforming DNA (776 aa),
fasta scores opt: 355 z-score: 368.7 E(): 3.8e-13 26.3%
identity in 635 aa overlap and Mycobacterium tuberculosis
TR:P71729(EMBL:Z81368) hypothetical 52.8 kD protein (514
aa), fasta scores opt: 450 z-score: 470.1 E(): 8.5e-19
35.2% identity in 546 aa overlap. Contains multiple
possible membrane spanning hydrophobic domains."
  /codon_start=1
  /transl_table=11
  /product="putative integral membrane protein."
  /protein_id="CAB66245.1"
  /db_xref="GI:6714673"
  /translation="MSRHLPGADRPDGLAGSRPPPPRAVHAASGHRLLGAHP
EGPADRLVPPALAAATAAVTLDPAGWAVGVTTGALLAGALLSASPRGRSRAPVA
ALLCVAAASAALHGDRLRRGPVPAARHYATVTAEVETGDPRLTRPRVGRDRAV
PPTVLEAEVRVSGAAGATVTTPTPLVLDVVDGGDSRATRSARAWLTLLPTRL
RYTAKLSPRRGRDRIAAVLRVRAGRPPELVIAEFAAQRLAGRLRGRELRATEDLPA
DARALLGVVGDTSRVTPLEAEKFTDLTHTLAVSGANFTVILLILGPPGLAQRS
ERRGLAPRLGLPTLTALLGLASFVYVCPDPSPVLRAAACGVALLATATGRRS
LLPALATVLLVLDVLPWLSYGLLSVLATGALLTLAPRMSAALRRRVPRLAE
LAAAGAAALCAPVAVI.SARVSLVPCNLIAEFAPATVGLFAALATAPAAAPLA
KSLAWCGGPAEWIAIARTGAALPGAGVDWPGWAGALLALVTVGLLAGRLLRH
PWCAGACALLILVTVQPPPLRLVLLKHPGGRWVNCVDGCDLVLAAGBGTGVV
DTGPDPAVLDHCLRALGTRVPLVLLTHFHADHAGVAGLVGRVTVGAIEETALEEPA
EQAEFVRQARARIPVRHAATGEQRSGPLSQWVYVPPSPAPTAFTCEGPNDA
VTLVYTAGMRVLLGLDLEPPAQRASLAPAAAALKDVLKVAHGSAAHODFAFLRR
IAPRLALISCCEDNSYGHGPACTVAALRAQATVLTDRSGSVAVAGRGLRVAGD"
  /complement(5757..6824)
  /gene="SCC123.06c"
  /complement(5757..6824)
  /gene="SCC123.06c"
  /note="SCC123.06c, possible DNA-binding protein, len: 355
aa. Similar to Mycobacterium tuberculosis
TR:P71728(EMBL:Z81368) hypothetical 30.5 kD protein (297
aa), opt: 271 z-score: 260.6 E(): 4e-07 36.3% identity
in 292 aa overlap. Also similar in parts to Bacillus

```


Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE

JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: LI0182
Center clone name: 46_L_10

*** NOTE: This record contains 66 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
708 807: contig of 707 bp in length
808 1505: contig of 698 bp in length
1506 1605: gap of 100 bp
1606 2303: contig of 698 bp in length
2304 2403: gap of 100 bp
2404 3111: contig of 708 bp in length
3112 3211: gap of 100 bp
3212 3930: contig of 719 bp in length
3931 4030: gap of 100 bp
4031 4731: contig of 701 bp in length
4732 4831: gap of 100 bp
4832 5543: contig of 712 bp in length
5544 5643: gap of 100 bp
5644 6337: contig of 694 bp in length
6338 6437: gap of 100 bp
6438 7148: contig of 711 bp in length
7149 7248: gap of 100 bp
7249 7946: contig of 698 bp in length
7947 8046: gap of 100 bp
8047 8738: contig of 692 bp in length
8739 8838: gap of 100 bp
8839 9524: contig of 686 bp in length
9525 9624: gap of 100 bp
9625 10323: contig of 699 bp in length
10324 10423: gap of 100 bp
10424 11110: contig of 687 bp in length
11111 11210: gap of 100 bp
11211 11910: contig of 700 bp in length
11911 12010: gap of 100 bp
12011 12704: contig of 694 bp in length
12705 12804: gap of 100 bp
12805 13492: contig of 688 bp in length
13493 13592: gap of 100 bp
13593 14282: contig of 690 bp in length
14283 14382: gap of 100 bp
14383 15078: contig of 696 bp in length
15079 15178: gap of 100 bp
15179 15881: contig of 703 bp in length

15882 15981: gap of 100 bp
15982 16683: contig of 702 bp in length
16684 16783: gap of 100 bp
16784 17504: contig of 721 bp in length
17505 17604: gap of 100 bp
17605 18298: contig of 694 bp in length
18299 18398: gap of 100 bp
18399 19087: contig of 689 bp in length
19088 19187: gap of 100 bp
19188 19871: contig of 684 bp in length
19872 19971: gap of 100 bp
19972 20661: contig of 690 bp in length
20662 20761: gap of 100 bp
20762 21462: contig of 701 bp in length
21463 21562: gap of 100 bp
21563 22265: contig of 703 bp in length
22266 22365: gap of 100 bp
22366 23065: contig of 700 bp in length
23066 23165: gap of 100 bp
23166 23873: contig of 708 bp in length
23874 23973: gap of 100 bp
23974 24676: contig of 703 bp in length
24677 24776: gap of 100 bp
24777 25507: contig of 731 bp in length
25508 25607: gap of 100 bp
25608 26297: contig of 690 bp in length
26298 26397: gap of 100 bp
26398 27083: contig of 686 bp in length
27084 27183: gap of 100 bp
27184 27879: contig of 696 bp in length
27880 27979: gap of 100 bp
27980 28674: contig of 695 bp in length
28675 28774: gap of 100 bp
28775 29461: contig of 687 bp in length
29462 29561: gap of 100 bp
29562 30243: contig of 682 bp in length
30244 30343: gap of 100 bp
30344 31064: contig of 721 bp in length
31065 31164: gap of 100 bp
31165 31860: contig of 696 bp in length
31861 31960: gap of 100 bp
31961 32669: contig of 709 bp in length
32670 32769: gap of 100 bp
32770 33480: contig of 711 bp in length
33481 33580: gap of 100 bp
33581 34283: contig of 703 bp in length
34284 34383: gap of 100 bp
34384 35083: contig of 700 bp in length
35084 35183: gap of 100 bp
35184 35907: contig of 724 bp in length
35908 36007: gap of 100 bp
36008 36729: contig of 722 bp in length
36730 36829: gap of 100 bp
36830 37539: contig of 710 bp in length
37540 37639: gap of 100 bp
37640 38334: contig of 695 bp in length
38335 38434: gap of 100 bp
38435 39155: contig of 721 bp in length
39156 39255: gap of 100 bp
39256 39955: contig of 700 bp in length
39956 40055: gap of 100 bp
40056 40792: contig of 737 bp in length
40793 40892: gap of 100 bp
40893 41591: contig of 699 bp in length
41592 41691: gap of 100 bp
41692 42411: contig of 720 bp in length
42412 42511: gap of 100 bp
42512 43231: contig of 720 bp in length
43232 43331: gap of 100 bp
43332 44040: contig of 709 bp in length
44041 44140: gap of 100 bp
44141 44840: contig of 700 bp in length
44841 44940: gap of 100 bp

```
* 44941 45640: contig of 700 bp in length
* 45641 45740: gap of 100 bp
* 45741 46442: contig of 702 bp in length
* 46443 46542: gap of 100 bp
* 46543 47250: contig of 708 bp in length
* 47251 47350: gap of 100 bp
* 47351 48061: contig of 711 bp in length
* 48062 48161: gap of 100 bp
* 48162 48853: contig of 692 bp in length
* 48854 48953: gap of 100 bp
* 48954 49680: contig of 727 bp in length
* 49681 49780: gap of 100 bp
* 49781 50484: contig of 704 bp in length
* 50485 50584: gap of 100 bp
* 50585 51294: contig of 710 bp in length
* 51295 51394: gap of 100 bp
* 51395 52092: contig of 698 bp in length
* 52093 52192: gap of 100 bp
* 52193 52884: contig of 692 bp in length.
FEATURES
    source          Location/Qualifiers
                    1..52884
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="RP11-46L10"
                        /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT      10147 a 13058 c 12120 g 10692 t 6867 others
ORIGIN

Query Match          9.78; Score 52.4; DB 73; Length 52884;
Best Local Similarity 49.1%; Pred. No. 1.7;
Matches 131; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 1 cccgagctgtgagggcgagccgggtataaagcctcgtggccttgcgcggcagccg 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40493 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 40552

QY 61 caggttcccgccgccccgagccccgcgcgcctgaagctgcgcgcctctctgggctct 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40553 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 40612

QY 121 gctgtgccccctctcgtcagctccgctgctgtcttcttagtggtgcgcaagctgtgg 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40613 GCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 40672

QY 181 cccagctgtcgtcgtcgtgagtcggtgcgcggagggcgggcgcccttgcccaacc 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40673 CCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 40732

QY 241 cctcggcacctcaaccgctgaac 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40733 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 40759

RESULT 11
LOCUS      AC090977 149719 bp DNA HTG 17-APR-2001
DEFINITION Mus musculus chromosome 16 clone rp23-213m14, WORKING DRAFT
            SEQUENCE, 40 unordered pieces.
ACCESSION  AC090977
VERSION     AC090977.6 GI:13654359
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 149719)
AUTHORS   Milam,J., Ford,B., Hine,R. and Roe,B.A.
TITLE     Mus musculus Chromosome 16 BAC Clone rp23-213m14
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 149719)
AUTHORS   Milam,J., Ford,B., Hine,R. and Roe,B.A.
TITLE     Direct Submission
JOURNAL   Submitted (22-MAR-2001) Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Apr 17, 2001 this sequence version replaced gi:13592223.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2175: contig of 2175 bp in length
2275: gap of unknown length
2276: contig of 2132 bp in length
4407: gap of unknown length
4408: contig of 2522 bp in length
7029: gap of unknown length
7123: gap of unknown length
9583: contig of 2454 bp in length
9683: gap of unknown length
11961: contig of 2278 bp in length
11962: gap of unknown length
12061: contig of 2221 bp in length
14282: gap of unknown length
14382: gap of unknown length
17116: contig of 2734 bp in length
17217: gap of unknown length
19196: contig of 1980 bp in length
19197: gap of unknown length
19296: gap of unknown length
21939: contig of 2643 bp in length
22040: gap of unknown length
22041: contig of 2673 bp in length
24812: gap of unknown length
24813: contig of 2623 bp in length
27436: gap of unknown length
27536: contig of 2554 bp in length
30090: gap of unknown length
30190: contig of 2574 bp in length
32863: gap of unknown length
34849: contig of 1986 bp in length
34950: gap of unknown length
37203: contig of 2354 bp in length
37404: gap of unknown length
39633: contig of 2230 bp in length
39734: gap of unknown length
43074: contig of 3340 bp in length
43174: gap of unknown length
45900: contig of 2727 bp in length
45901: gap of unknown length
46001: contig of 3294 bp in length
49295: gap of unknown length
49395: contig of 3531 bp in length
52926: gap of unknown length
53026: contig of 2365 bp in length
55391: gap of unknown length
55491: contig of 3728 bp in length
59219: gap of unknown length
59319: contig of 4147 bp in length
63466: gap of unknown length
63566: contig of 1984 bp in length
65549: gap of unknown length
65550: gap of unknown length
69789: contig of 4140 bp in length
69889: gap of unknown length
72576: contig of 2687 bp in length
72577: gap of unknown length
72677: contig of 3205 bp in length
75881: gap of unknown length
75882: contig of 3779 bp in length
79860: gap of unknown length
84650: contig of 4790 bp in length
84750: gap of unknown length
88672: contig of 3922 bp in length
88673: gap of unknown length
92936: contig of 4163 bp in length
93035: gap of unknown length
```



```

DHTEPREMYTFHPVAGORGNVAVQADTVGGAVPIAVNVNTENGSTTQVTVATPPGAAN
GRALAYLVHRTGDTCTGLTVDVSTIFILFAGASAMAGLNIPIVRLPRYGMADPWAR
ATPLVLVHTFGIAALVTLVFOANVDAQAGAYATGVLALMTSAAILVFLTELRRKHQA
ALAGLVSLFVTVTAFTVTKERPEGLWIALCEIAGILVYSINRIGRSTELRVQNVHF
DEAGLDWVRDVAAGLPLVRIHNLNAGDAEYQERALEVRLDNHLHSAGAEVAFLEVA
VLDAOSKSTVILEGVQVPHRILRLATGNSVPNTIAAVLLYVRDLTGAPPHYFEWSE
KGPAAALRFLLAGEGDIPPLTHEILRAERDPRRQVHVGG"
/gene="DR2266"
/complement(4130..4459)
/gene="DR2266"
/complement(4130..4459)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFL1820.1"
/db_xref="GI:6460080"
/translation="MRAVAAAAYKLELNRKNVMEQWSLRDLYRTLDMPKNPLRDA
QERLDAVSAAYGLPAGADMDFLLALNAXVAAAEARGAAVTGGLPAGLNTADFVTA
DAVRPLG"
/complement(4478..7000)
/gene="DR2267"
/complement(4478..7000)
/notes="similar to GB:AL009126 percent identity: 50.81;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA modification methyltransferase-related
protein"
/protein_id="AAFL1813.1"
/db_xref="GI:6460073"
/translation="MPQTEAORMEDFVAYWRTLKGDKESOVFLDRLPQAFGHAGY
KEAAELEYVAKQGGKKEADLLWRPVLIEMKRCEKLANHYQQAQFYMLKLVDR
PRVAVLCNDFELWYDFNQQLDEPMDRLRIEELPERVTVLNFMEQERAPLFNNRVD
VREAAADSAVKLVNSVARGEDRARQAFLLQCVMAFAEDFELIPRGFTLADDAAR
AGRSFDFLGGFDFROMNTSERAGRFAPIPYNGGLFRAPDPIELNRDELYLLHKA
ALENNAKAIQPIFGVLFSQMDKQKHAKGHAYTSEADIMRVVLPITVTPFORQIEA
ATQKELRALDLEASFOVLDPACGSGNFYVAYRELRLEARALLRLDLSPAGTAL
PPARVSIROMHGLXPDPGVELAKVTLTAKELAIEMHDLGNTGLDPOPLDNL
DRIVOGDALFTPHWRDALVGNPFPQSKNKLQREMGAAVYKLRHAPYPRADYC
VYWRKAHODSGRAGLVGTLNTRKQNSDGLDLYVQHGHTITDAVGTQWSDGA
AHVSVINWVKGPAEGKHLAWQGHRTSPWQSTELPINSALSAGTQVTOAKLRV
NMNSGACYOQGTGKFLDGLGAEQMSAEKNAEVIPIYLTGDELLTSPHPHTR
YVIDFQFQVFGARAYKLPARIEREVLPTROAAAEEEAENAEVLAANPKAKTNKHH
RNFNLQWALYSGRSEMIKISLSRYIVCSRVTKQVFPFLDNGIRPDSGLQIFAFE
DYDSFGVQSSVHWQLIARGSLTLARLWYTSVDTPFPWP"
/notes="DR2268"
/notes="DR2268"
/notes="similar to PID:1651949 percent identity: 64.49;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="aminotransferase, class I"
/protein_id="AAFL1814.1"
/db_xref="GI:6460074"
/translation="MNASORASVPGSVFSLMDAAKMRARGRGLNVTDLISGSDQHP
PLSVDELRAATHDPATYRLLPFLSDTQPLRDAAYALARRFGVRLNADTEVLPLIGAQ
EGLAHLALLVATPDGDTLLPLDPPCPYPYWGAAVAGLVNVTPLPLPERGLPLDSRVA
NVRPRLLLNPNPNTSIAIAEFFREAAWCREGTLVLVHDHPYALTEFGDYPAPSA
LEAGMEGVVHLSLSTHHLGFRVGFVAGDAGAAALARVKAIDIPHVYLGITORAAA
HAILGLDEVRGVGARIFEERRDALLSALRDLGWVOTPOASMYAVACVPLGRDSVAPA
VRAAEITGVVSPQAFGERGEFVRFALVQPEVRLRGAALKGEVGV"
/complement(8295..9002)
/gene="DR2269"
/complement(8295..9002)
/notes="similar to GB:M21151 SP:P12282 PID:145540 GB:U00096
PID:1651375 percent identity: 59.59; identified by
sequence similarity; putative"
/codon_start=1

```

```

/transl_table=11
/product="molybdopterin biosynthesis MoeB"
/protein_id="AAFL1815.1"
/db_xref="GI:6460075"
/translation="MTLREELRRVSRPLVPWEAAGAOKELRAARLVVVGAGGLGG
PVTRQAGAGVAGTADGTVSTVNLHQOLYRTADVGRSKAETACAAVQGNPFVQ
IQAAPALTPTETATLIAAHLTVADTNFETRYLIADTRELGRWWGASVSGLIC
SVFQPHGLRDFVPTPEDDSSCDELGVLPVPLVGLDMALQALLVLGGVGEPLRGKL
WTFDALTGVRVRLKMG"
/notes="DR2270"
/notes="DR2270"
/notes="similar to GB:AL009126 percent identity: 54.96;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFL1816.1"
/db_xref="GI:6460076"
/translation="MRAVVSAAVVFVLSVLGSLVAGVLSRGRGEDIIRGDLPGGSG
TYROYKGAAAVTLADIILKGAAGVGLALWLAQALPLATATATFGVFGHGVVWFG
FROGGGAPFLGAMLVVAPWTLATVTFALALPLIYRATLQPLRLNLAIFATVAVVP
VGLLIASRLGSGAEFLAGSAAAMGIRAVHLLAEQRA"
/notes="DR2271"
/notes="DR2271"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFL1821.1"
/db_xref="GI:6460081"
/translation="MKRTLALLALAGTASAEVRVDGQTVRVLSGQVTVQWTFAPA
LGVSVRVLNDQNVAVGVVYALSDGSRVLGRADLPGPVTSLDSSGGAVRVTVGGA
GRSERLTGDTGALPEVRVFTPDPAVTGWLARFADAPALFRAQAQFPAFPF
IALREAAAKTRGDYDAAALSARVRLGVSPFPAGTQALAAALDRAGPAAANLALDRA
KDAARAGLDPAPVSRALFAYGNPSGYVTLLAONRLPRAEVMFSLRELYPRFEG
GDALYRYADVLDQAQGRAGEEWRQFARSRLBETLYNLGPNDRVLRDVARLATFAL
Query Match 9.6%; Score 52; DB 1; Length 12924;
Best Local Similarity 46.1%; Pred. No. 3.7;
Matches 175; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
Qy 5 cgcctggagggcgagaccgggtataagaagcctcgtagcctcccgccgagcgagcgag 64
Db 8766 CGCCGATGTGGGCGCTCCAAAGCGGAACCGCTGCGCCCGTCCGCCAGGGGGTCAATCC 8707
Qy 65 tccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 124
Db 8706 GTTCGTGCAGATCAGCGCGCGCCAGCGCTGACGCGGAGAGACCGCCCGCATCGC 8647
Qy 125 ggcctgtcctgcagctccgctgctgtcttcttagtggtcgccagagcctgtggtccca 184
Db 8646 CGCGCACACCTGACGGTGGACGCCACCGACAACCTTTGAAACCCCGCTACCTCATCGCGGA 8587
Qy 185 gctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 244
Db 8586 CACTGCGCGGAACCTGGCGCGCGAGTGGGTGTGGGGCGCTGCCAGCGGGGTACGGGGCT 8527
Qy 245 cggcacctcaaccgcgtgaagctcctctgtagcagcctgggcatcccccgtgaaccacct 304
Db 8526 GTGCAGCGCTTCGCGCGCGCACTCGGCTCGCGCACGTGTTTCCACACCCCGGAGACGA 8467
Qy 305 catagagggctcccccagaagtgtgtgctgagctgggtcccccagagccgtggtggggccgtgaa 364
Db 8466 CAGCTCTGTGACGAACCTCGGCGCGTCTCGGCGCGGTGCGCCCAACCTCGTGTGGCGACCTGAT 8407
Qy 365 ggcctgaagccctgctgg 384
Db 8406 GGCGCTCAAGCCCTGCTCG 8387

```

```
RESULT 13
AC011352
LOCUS
DEFINITION
AC011352 160949 bp DNA HTG 05-MAY-2000
Homo sapiens chromosome 5 clone CTC-327F10, WORKING DRAFT SEQUENCE,
3 unordered pieces.
AC011352.3 GI:7710577
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (06-Oct-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 5, 2000 this sequence version replaced gi:6604445.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 310877, H284
Center clone name: CIT-HSPC_327F10
-----
Summary Statistics
Consensus quality: 147081 bases at least Q40
Consensus quality: 155898 bases at least Q30
Consensus quality: 158709 bases at least Q20
Estimated insert size: 165000; pulse field gel estimation
Estimated insert size: 160749; sum-of-contigs estimation
Quality coverage: 5.95 in Q20 bases; pulse field gel estimation
Quality coverage: 6.1 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9538: contig of 9538 bp in length
* 9539 9638: gap of unknown length
* 9639 68284: contig of 58646 bp in length
* 68285 68384: gap of unknown length
* 68385 160949: contig of 92565 bp in length.
Location/Qualifiers
1. .160949
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-327F10"
/clone_lib="Caltech human BAC library C"
50420 a 29821 c 31200 g 49306 t 202 others
BASE COUNT
ORIGIN

Query Match
Best Local Similarity 9.4%; Score 51.2; DB 62; Length 160949;
Matches 77; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 256 acscgctgaagctctgtcagcagcctgggcatcccttgagaccacctatagaggct 315
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18608 ATCCATTAAAGCTTCCTCGAAACCTCTGGGCACTTCTGTGAGCACCCTGTGGAGGGC 18667
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 316 cc-agaagtgtgtgctgagctgggtcccccagggccgtg9ggggccgtgaaggccctgaagg 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome Mol. Microbiol. 21 (1), 77-96 (1996)									
JOURNAL	97000351									
MEDLINE	2 (bases 1 to 37931)									
REFERENCE	Brown, S.P. and Harris, D.									
AUTHORS	Unpublished									
JOURNAL	3 (bases 1 to 37931)									
REFERENCE	Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.									
AUTHORS	Submitted (28-JUN-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge									
JOURNAL	CB10 1SA E-mail: barrllesanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK									
COMMENT	Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/jun/cgi-bin/frameplot.pl . CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid D10 .									
FEATURES	Location/Qualifiers 1. .37931 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid D10" 1. .786 /gene="SCD10.01" /note="SCD10.01" /note="SCD10.01" content in glycine and aspartate amino acid residues. Contains 2x Pfam match to entry PF01839 FG-GAP, FG-GAP repeat /codon_start=1 /transl_table=1 /product="hypothetical protein SCD10.01c (fragment)" /protein_id="CAB95880.1" /db_xref="GI:8894719"									
gene										
CDS										
misc_feature	1. .498 /gene="SCD10.01"									

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:48:23 ; Search time 115.94 Seconds
(without alignments)
1060.700 Million cell updates/sec

Title: US-09-700-770-6

Perfect score: 543

Sequence: 1 ccggcgctggaggggcgagg.....gttaagagcaaaaaaaaaa 543

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_NA:*
- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
 - 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
 - 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
 - 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
 - 5: /cgn2_6/ptodata/1/ina/PCRUS_COMB.seq.*
 - 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	56.4	10.4	263	2	US-08-964-725-2
2	56.4	10.4	507	2	US-08-964-725-4
3	56.4	10.4	519	2	US-08-964-725-5
4	45.6	8.4	6453	1	US-08-306-691B-14
5	45.6	8.4	6453	3	US-09-356-952-8
6	45.4	8.4	2721	6	5215881-2
7	45.4	8.4	8438	1	US-07-945-283-1
8	44.4	8.2	1209	4	US-09-105-537-21
9	44.4	8.2	13613	4	US-09-105-537-3
10	44.4	8.2	38506	3	US-09-320-878-19
11	44	8.1	6453	3	US-09-209-668-10
12	43.6	8.0	432	1	US-08-642-255-48
13	43.6	8.0	756	1	US-08-642-255-50
14	42.8	7.9	13842	4	US-09-105-537-30
15	42.8	7.9	36778	4	US-09-105-537-5
16	42.6	7.8	2543	3	US-08-555-669-11
17	42.6	7.8	2543	3	US-09-073-663-11
18	42.2	7.8	30001	2	US-08-125-468-1
19	42.2	7.8	30001	2	US-08-474-933-1
20	42.2	7.8	4403765	4	US-09-103-840A-2
21	41.8	7.7	4257	2	US-08-690-473-1
22	41.8	7.7	4257	4	US-09-259-821A-1
23	41.8	7.7	4257	4	US-08-843-659-1
24	41.8	7.7	12001	1	US-08-458-568A-11
25	41.6	7.7	530	3	US-08-758-662-4
26	41.6	7.7	4085	1	US-08-486-270-7
27	41.6	7.7	4085	3	US-08-367-264-7

28	41.6	7.7	4181	1	US-08-486-270-9
29	41.6	7.7	4181	3	US-08-367-264-9
30	41.6	7.7	4524	2	US-08-845-998-7
31	41.6	7.7	4524	2	US-08-845-998-7
32	41.6	7.7	4524	3	US-09-206-537-7
33	41.6	7.7	4524	3	US-09-206-537-7
34	41.6	7.7	4524	4	US-09-430-854-7
35	41.6	7.7	4524	4	US-09-430-854-7
36	40.6	7.5	2003	1	US-08-459-526A-21
37	40.6	7.5	2003	2	US-08-734-591A-21
38	40.6	7.5	2003	3	US-08-341-018-71
39	40.6	7.5	2003	3	US-08-470-335-21
40	40.6	7.5	2003	4	US-08-735-021-21
41	40.6	7.5	2003	4	US-08-734-664A-21
42	40.6	7.5	2003	4	US-08-470-339-21
43	40.4	7.4	1565	3	US-09-320-878-24
44	40.4	7.4	4195	1	US-08-340-011-1
45	40.4	7.4	4195	3	US-08-901-710-1

ALIGNMENTS

RESULT 1
US-08-964-725-2
; Sequence 2, Application US/08964725
; Patent No. 5939265
; GENERAL INFORMATION:
; APPLICANT: COHEN, Maurice
; APPLICANT: FRIEDMAN, Paula N.
; APPLICANT: GORDON, Julian
; APPLICANT: HODGES, Steven C.
; APPLICANT: KLASS, Michael R.
; APPLICANT: KRATOCHVIL, Jon D.
; APPLICANT: ROBERTS-RAPP, Lisa
; APPLICANT: RUSSELL, John C.
; APPLICANT: STROUPE, Steven D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; TITLE OF INVENTION: FOR DETECTING DISEASES OF THE LUNG
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,725
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 5997.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

RESULT

```

US-09-356-952-8
; Sequence 8, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Boriack-Sjodin, Ann
; APPLICANT: Margarit, S. M.
; APPLICANT: Bor-Sogi, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 600-1-228N
; CURRENT APPLICATION NUMBER: US/09/356,952
; CURRENT FILING DATE: 1999-07-19
; EARLIER APPLICATION NUMBER: 60/093,631
; EARLIER FILING DATE: 1998-07-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0

```

```

? : BORINGMO. Valencia ver. 2.0
? : SEQ ID NO 8
? : LENGTH: 6453
? : TYPE: DNA
? : ORGANISM: Homo sapiens
US-09-356-952-8

Query Match      8.4%      Score 45.6;   DB 3;   Length 6453;
Best Local Similarity 47.8%
Matches 132; Conservative
0; Mismatches 144; Indels 0; Gaps 0;

```

Qy	229	ccctggccacccccctcctggcacccctcaacccgcctgaagctcctctgtgacgagctctgggca	288
Db	2604	caatgacctctcctctgcacgagcgacctgtgctgcttagctaccagctccgggacc	2663
Qy	289	tcacctgaaaccacctcatagaggctcccaagaagtgtgtgctgagctgggtcccag	348
Db	2664	ctctgggacccccgggacccatgtgaccacgcgccccctgcactgtagctctccggg	2723
Qy	349	ccgtggggccgtggaaggccctgaaaggccctgctggggccctgacagtgtttgctcag	408
Db	2724	acggcaggcgagtgaggagggcgagggcggggtctgggctcaagccctgcagctcctgg	2783
Qy	409	ccaagactgagcatctacacttgagacagacagactgccaccgcagggcctgaaac	468
Db	2784	ccgacacagctccgggaaggcgaggtctcttggggagagctgcctctgagccagcccgga	2843
Qy	469	ccgcgcgggggagacgtccatcccccttcccgc	504
Db	2844	acagtaccctgaagcccaacccctcttaccacg	2879

db 2844 acggtgacccctggggcccccctcttgcctccca 2879

RESULT 6
5215881-2/c
; Patent No. 5215881
; APPLICANT: CHEUNG, ANDREW K.
; TITLE OF INVENTION: PSEUDORABIES DIAGNOSIS PROBLEMS
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/537,855
; FILING DATE: 13-JUN-1990
; FILING DATE:


```

; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match      8.2%; Score 44.4; DB 3; Length 38506;
Best Local Similarity 46.7%; Pred. No. 0.38;
Matches 141; Conservative 0; Mismatches 161; Indels 0; Gaps

QY 1 ccggcgctggaggcgaggaccgggtataagaagcctctggccttgcccgggcgagccg 60
    || || || || || || || || || || || || || || || || || || || ||
Db 35677 ccggggtgtccgcggggagacgagctggcgggccggcggtcgccggcgggcgcatg 35736

QY 61 caggttcccccgcgcccgcgagcccccgccatgaagctgcgcctcctctgggctct 120
    || || || || || || || || || || || || || || || || || || || ||
Db 35737 tctgtctgaccgcgcgacggcgccgggagccggaggtcttcacgagaccggagcgt 35796

QY 121 gcgtggccctctctgcagctccgctgctgtcttcttagtgaggctcggccaaagcctgtg 180
    || || || || || || || || || || || || || || || || || || || ||
Db 35797 tcgacctgcgcgccccgcagcgcgcgcgcgcacctcgcgctcgaccccgcggtcccgta 35856

QY 181 ccagcctgtcgtgcgtggagtcggcgcgaggcgccggggcgccgagacctggccaacc 240
    || || || || || || || || || || || || || || || || || || || ||
Db 35857 gccgggtggctccctgttcgggttcaggcgaggtcgcgctcgagacctggccgggc 35916

QY 241 cctcggcaccctcaaccgcgtgaagctcctgctgagcagcctggggcctccccgtgaacc 300
    || || || || || || || || || || || || || || || || || || || ||
Db 35917 gtttcccggtgtgcggcaggcggggacgtgctccgcccccgcgcgctgtcgcc 35976

QY 301 ac 302
    |
Db 35977 gc 35978

; RESULT 11
US-09-209-668-10
; Sequence 10, Application US/09209668A
; Patent No. 6114517
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Xu, Xiaoxing S.
; TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR
; TITLE OF INVENTION: alpha-INDUCED EXPRESSION OF CELL ADHESION MOLECULES
; FILE REFERENCE: ISPH-0336
; CURRENT APPLICATION NUMBER: US/09/209,668A
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 6453
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1664)..(1774)
; FEATURE:
; NAME/KEY: CDS

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 13:34:21 ; Search time 2570.26 Seconds
(without alignments)
1997.035 Million cell updates/sec

Title: US-09-700-770-6
Perfect score: 543
Sequence: 1 ccgcgcgtgaggcgagg.....gttaagagcaaaaaaaaaa 543

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_estl1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estov1:*
70: em_estov2:*
71: em_estov3:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.


```

QY 471 cgccgggggagaccgtccatccctcccccggccctctcaataaaacgtgtaaga 530
|||||
Db 65 CGCCGGGGGAGGACCGTCCATCCCTTCCCGCGGCCCTCTCAATAAACGTGGTTAAGA 6
|||||
QY 531 gcaaa 535
|||||
Db 5 GCAA 1

RESULT 4
AI333740/c
LOCUS      424 bp      mRNA      EST      28-DEC-1998
DEFINITION gp98f05.x1 Soares_fetal_lung_NbHL19w Homo sapiens cDNA clone
IMAGE:1931073 3', mRNA sequence.
ACCESSION  AI333740
VERSION     AI333740
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 424)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 413.

FEATURES             Location/Qualifiers
     source           1..424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Soares_fetal_lung_NbHL19w"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT73D (Pharmacia) with a
                     modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGCGCAATTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT73 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHL19w."
BASE COUNT      70 a 139 c 149 g 66 t
ORIGIN

Query Match      78.1%; Score 424; DB 19; Length 424;
Best Local Similarity 100.0%; Pred. No. 1.9e-81;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 ctggggctctgctggccctgtctcgcagctcccgctgctgtctttctagtggcgccgccc 170
|||||
Db 424 CTGGGGCTCTGCGTGGCCCTGTCTCGAGCTCCGCTGCTGCTTTCTTAGTGGGCTCGGCC 365
|||||
QY 171 aagcctgtgcccagcctgtctgctgctgctgaggtcgcgcgaggccggcgccgagcc 230
|||||
Db 364 AAGCCTGTGGCCCAAGCCTGTGCTGCTGCGCTGAGTTCGGCGGAGCGGGCGCGGAGCC 305
|||||
QY 231 ctggcccaacccctcgacacccctcaacccctcctcctcctgctgagcagcctggcctc 290
|||||
Db 304 CTGGCCAAACCCCTCGGCAACCCCTCAACCCCTGAAGCTCTCTGCTGAGCAGCCTGGGCATC 245
|||||
QY 291 cccgtgaaccacacctatagagggtcccccagaagtgtgtggtgagctgggtcccccaggcc 350
|||||

```

```

Db 244 CCCGTGAACCACTCATAGAGGCTCCAGAAAGTGTGTGGCTGAGCTGGGTCCCGAGGCC 185
|||||
QY 351 gtgggggcccgtgaagccctgaagccctgctggggccctgacagtgatttggtgagcc 410
|||||
Db 184 GTGGGGCCCTGAAGGCCCTGAAGCCCTGCTGGGGGCCCTGACAGTGTGTTGGGTGAGCC 125
|||||
QY 411 gagactgagcatctacacctgagagacaagacgctgccaccgccgagggctgaaaaccc 470
|||||
Db 124 GAGACTGGAGCATCTACACCTGAGACAAGACGCTGCCACCCCGCGAGGGCTGAAAACCC 65
|||||
QY 471 cgccgggggagaccgtccatccctcccccggccctctcaataaaacgtgtaaga 530
|||||
Db 64 CGCCGGGGGAGGACCGTCCATCCCTTCCCGCGGCCCTCTCAATAAACGTGGTTAAGA 5
|||||
QY 531 gcaa 534
|||||
Db 4 GCAA 1

RESULT 5
AI744099/c
LOCUS      424 bp      mRNA      EST      17-DEC-1999
DEFINITION wc36c10.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2317266 3',
IMAGE sequence.
ACCESSION  AI744099
VERSION     AI744099
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 424)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert length: 495 Std Error: 0.00
            Seq primer: -40UP from Gibco.

FEATURES             Location/Qualifiers
     source           1..424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_image="IMAGE:2317266"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clonelids
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      68 a 133 c 146 g 77 t
ORIGIN

Query Match      77.5%; Score 420.8; DB 24; Length 424;
Best Local Similarity 99.5%; Pred. No. 9.4e-81;

```


Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bhrp/image/image.html
Insert length: 497 Std Error: 0.00
Seq primer: -400p from Gibco.

FEATURES

source
1. .413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2317080"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"

/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 68 a 135 c 146 g 64 t

ORIGIN

Query Match 76.1%; Score 413; DB 24; Length 413;

Best Local Similarity 100.0%; Pred. No. 4.4e-79;
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 tgcgtggccctgtcctgcagctccgctgctcttcttagtggcgcgccaagcctgtg 179

Db 413 TGCCTGGCCCTGTCTGCAGCTCCGCTGCTTTCTTAGTGGCTCGGCCAAGCCTGTG 354

Qy 180 gccagcctgtcgtgcgtgagtcgctgcgcgagcgccggcgccgacctggccaac 239

Db 353 GCCCAGCCTGTGCTGGCTGGAGTCGGCGCGGAGCGCGGGCGGACCTGGCCAAAC 294

Qy 240 cccctcgacacctcaaccgctgaagctcctgctgagcagcctggcgcatcccgctgaac 299

Db 293 CCCCCTGGCACCTCAACCCGCTGAAGCTCCTGCTGAGCAGCCTGGGATCCCGTGAAC 234

Qy 300 caccctatagagggtccacagaagtgtgtggtgagctgggttccccagggcctggggccc 359

Db 233 CACCTCATAGAGGGCTCCAGAAAGTGTGTGCTGAGCTGGGTCCGCCAGCGCTGGGGCC 174

Qy 360 gtgaagccctgaagccctgctggggccctgacagtgtttgctgagcagcagactgga 419

Db 173 GTGAAGGCCCTGAAGGCCCTCTGGGGCCCTGACAGTGTGTGCTGAGCGAGACTGGA 114

Qy 420 gcatctacacctgaggacaagacgtgcccaccgcgagggcgtgaaaaccgcgcggg 479

Db 113 GCATCTACACCTGAGGACAAGACGCTGCCACCGCGAGGGCTGAAACCCCGCGCGG 54

Qy 480 gagacgctccatcccttcccccgccctctcaataaactgtggttaagagc 532

Db 53 GAGGACCGTCCATCCCTTCCCGCGGCCCTCTCAATAAAGCTGTTAAGAGC 1

RESULT 8

BF221778/c 416 bp mRNA EST 09-NOV-2000
LOCUS 7062e02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3578595 3'
DEFINITION similar to TR:Q9VI31 Q9VI31 CG10690 PROTEIN. ;, mRNA sequence.
ACCESSION BF221778
VERSION BF221778.1 GI:11128955
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (Bases 1 to 416)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov

High quality sequence stop: 383.

FEATURES

Location/Qualifiers

source

1. .416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3578595"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"

/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 69 a 134 c 144 g 69 t

ORIGIN

Query Match 76.0%; Score 412.8; DB 146; Length 416;

Best Local Similarity 99.5%; Pred. No. 4.9e-79;
Matches 414; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 122 cgtggccctgtcctgcagctccgctgcttcttagtggcgcgccaagcctgtgac 181

Db 416 CGTGGCCCTGTCTGCAGCTCCGCTGCTTCTTAGTGGCTCGGCCAAGCCTGTGGC 357

Qy 182 ccagcctgtcgtgcgtggagtcgctgcgcgagcgccggcgccgacctggccaaccc 241

Db 356 CCAGCCTGTCTGCTGGCTGGAGTCGGCGCGGAGCGCGGGCGGACCTGGCCAACC 297

Qy 242 cctcggcaacctcaaccgcctgaagctcctctgagcagcctgggcatccccgtgaacca 301

Db 296 CCTCGCACCTCAACCGTTGAAGCTCCTTGTAGCAGCCTGGGCAATCCCGTGAACA 237

Qy 302 cctcatagagcctcccaagaagtgtgtgctgagctaggtcccccagccgtggggccgt 361

Db 236 CCTCATAGAGGGCTCCCAAGAAGTGTGTGCTGAGCTGGGTCCCCAGGCCGTGGGGCCGT 177

Qy 362 gaagggcctgaagccctgtcgtggggccctgacagtgtttggctgagcagagactggagc 421

Db 176 GAAGGCCCTGAAGGCCCTGTCTGGGGCCCTGACAGTGTTTGGCTGAGCCGAGACTGGAGC 117

Qy 422 atctacacctgagagacagcgtgccaccgcgcgagggctgaaacccccccgcgggga 481

Db 116 ATCTACACCTGAGGACAAGACGCTGCCACCCCGAGGGCTGAAACCCCGCCGCGGGA 57

Qy 482 ggaacgctcatcccttcccccgccctctcaataaactgtggttaagagcaaaaa 537

Db 56 GGACCGTCCATCCCTTCCCCCGGCCCTCTCAATAAAGCTGGTAAAGAGCAAAA 1

Db	198	GGTCCACAGCCGTGGGGCCGTGAAGCCCTGAAGCCCTGCTGGGGCCCTCACAGTG	139
Qy	399	tttggctgagcgcgagactgagcatctacacctgaggaacaagacgctgcccaccgcgag	458
Db	138	TTTGGCTGAGCGAGACTGCAGCATCTACACTGAGGACAGAGCGTGCACCCCGCGAG	79
Qy	459	ggctgaacaccccgccgggggagaccgtccatccccctccccggccctctcaataa	518
Db	78	GGCTGAAACCCCGCGCGGGGAGACCGTCCATCCCTTCGCCGCCCTCTCAATAA	19
Qy	519	acgtggttaagagcaaaa	536
Db	18	ACGTGTTAAGACAAA	1
RESULT	11		
AA649864/c			
LOCUS	AA649864	343 bp	mRNA
DEFINITION	ns54a03.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1187404	EST	13-NOV-1997
ACCESSION	AA649864		mRNA sequence.
VERSION	AA649864.1	GI:25771192	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 343)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-r@mail.nih.gov		
	Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.		
	cDNA Library Preparation: M. Bento Soares, Ph.D.		
	cDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html		
	Insert length: 539 Std Error: 0.00		
	Seq primer: -40m13 fwd. ET from Amersham		
	High quality sequence stop: 315.		
FEATURES	Location/Qualifiers		
source	1..343		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:1187404"		
	/clone_lib="NCI_CGAP_Pr22"		
	/sex="male"		
	/tissue_type="normal prostate"		
	/lab_host="DH10B"		
	/note="Organ: prostate; Vector: pT73p-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	48 a 111 c 119 g 65 t		
ORIGIN			
Query Match	62.6%;	Score 339.8;	DB 10; Length 343;
Best Local Similarity	99.4%;	Pred. No. 2.4e-63;	
Matches 341; Conservative	0;	Mismatches 2;	Indels 0; Gaps 0
Qy	195	gcgctgagtcgagcgcgagcgcggcgagccctggccacccctcgaccctc	254

Search completed: November 17, 2001, 13:34:25
Job time: 3424 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:53:48 ; Search time 4216.79 Seconds
(without alignments)
7556.355 Million cell updates/sec

Title: us-09-700-770-3
Perfect score: 2060
Sequence: 1 cttgagagctctcaataact.....ttccattgaaaaaaaaa 2060

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_bal.*
- 2: gb_bal2.*
- 3: gb_bal3.*
- 4: gb_in1.*
- 5: gb_in2.*
- 6: gb_in3.*
- 7: gb_om.*
- 8: gb_ov.*
- 9: gb_pat1.*
- 10: gb_pat2.*
- 11: gb_ph.*
- 12: gb_pl1.*
- 13: gb_pl2.*
- 14: gb_pl3.*
- 15: gb_pl4.*
- 16: gb_bal.*
- 17: em_bal2.*
- 18: em_fun.*
- 19: em_htgo_hum.*
- 20: em_htgo_inv.*
- 21: em_htgo_rod.*
- 22: em_htg_hum1.*
- 23: em_htg_hum2.*
- 24: em_htg_hum3.*
- 25: em_htg_hum4.*
- 26: em_htg_hum5.*
- 27: em_htg_hum6.*
- 28: em_htg_hum7.*
- 29: em_htg_hum8.*
- 30: em_htg_inv1.*
- 31: em_htg_inv2.*
- 32: em_htg_other.*
- 33: em_htg_rod.*
- 34: em_hum1.*
- 35: em_hum2.*
- 36: em_hum3.*
- 37: em_hum4.*
- 38: em_hum5.*
- 39: em_hum6.*
- 40: em_hum7.*
- 41: em_in.*
- 42: em_om.*
- 43: em_or.*

- 44: em_ov.*
- 45: em_pat.*
- 46: em_ph.*
- 47: em_pl.*
- 48: em_ro.*
- 49: em_sts.*
- 50: em_sy.*
- 51: em_un.*
- 52: em_vi.*
- 53: gb_sts1.*
- 54: gb_sts2.*
- 55: gb_sts3.*
- 56: gb_sy.*
- 57: gb_un.*
- 58: gb_v11.*
- 59: gb_v12.*
- 60: gb_htg1.*
- 61: gb_htg2.*
- 62: gb_htg3.*
- 63: gb_htg4.*
- 64: gb_htg5.*
- 65: gb_htg6.*
- 66: gb_htg7.*
- 67: gb_htg8.*
- 68: gb_htg9.*
- 69: gb_htg10.*
- 70: gb_htg11.*
- 71: gb_htg12.*
- 72: gb_htg13.*
- 73: gb_htg14.*
- 74: gb_htg15.*
- 75: gb_htg16.*
- 76: gb_htg17.*
- 77: gb_htg18.*
- 78: gb_htg19.*
- 79: gb_htg20.*
- 80: gb_htg21.*
- 81: gb_htg22.*
- 82: gb_htg23.*
- 83: gb_htg24.*
- 84: gb_htg25.*
- 85: gb_pr1.*
- 86: gb_pr2.*
- 87: gb_pr3.*
- 88: gb_pr4.*
- 89: gb_pr5.*
- 90: gb_pr6.*
- 91: gb_pr7.*
- 92: gb_pr8.*
- 93: gb_pr9.*
- 94: gb_ro1.*
- 95: gb_ro2.*
- 96: gb_in4.*
- 97: gb_pr10.*
- 98: em_bal3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1358.4	65.9	1438	88	AF098484	AF098484 Homo sapi
2	1351.8	65.6	1358	89	AF200345	AF200345 Homo sapi
3	1350.4	65.6	1519	10	E33553	E33553 ASP5. 2/200
4	1350.4	65.6	1648	10	E33552	E33552 ASP5. 2/200
5	1345.2	65.3	1365	88	AF090386	AF090386 Homo sapi
6	1199.2	58.2	1299	9	AR016588	AR016588 Sequence
7	1199.2	58.2	1299	9	AR078758	AR078758 Sequence
8	1151.8	55.9	1375	9	A70198	A70198 Sequence 2

```
9 1151.8 55.9 1375 9 A70213
10 1140.4 55.4 1350 88 AF090387 Homo sapi
11 1140.4 55.4 1352 89 AF200344 Homo sapi
12 1140.4 55.4 1353 88 AF098485 Homo sapi
13 949.2 46.1 1347 9 A70212
14 791.2 38.7 1360 9 AX046346
15 791.2 38.7 1360 9 AX046376
16 791.2 38.5 1479 94 D88899 Mouse mRNA
17 791.4 38.4 1520 94 AB038144 Mus muscu
18 788.8 38.3 1331 95 RNO251299 Rattus no
19 678.4 32.9 1840 10 AX086222 Sequence
20 678.4 32.9 1840 93 HSM801634
21 673.6 32.7 1804 93 HSA276704
22 397.2 19.3 481 9 A70208
23 396.4 19.2 1448 8 S49650
24 374.4 18.2 1647 88 AF078843
25 368 17.9 1191 8 AF312364
26 360.2 17.5 474 9 A70207
27 352.8 17.1 1442 8 PSI9838
28 343.6 16.6 1847 8 OM90321
29 331.2 16.2 1637 8 CHAJ7878
30 323.6 15.7 402 9 A70206
31 315.8 15.5 1708 8 DRE278268
32 315.4 15.5 59264 65 AC018278
33 315.4 15.5 127035 4 AC006574
34 315.4 15.5 272605 5 AC003669
35 316.6 15.4 2038 9 AK080625
36 316.6 15.4 2038 97 HMC2THD
37 315 15.3 1988 93 HSCATDC
38 313.8 15.2 1578 7 AB055312
39 303.4 15.0 377 9 A70210
40 308.4 15.0 123147 87 AC008655
41 308.4 15.0 198316 65 AC019157
42 307.8 14.9 2124 9 AX014348
43 304 14.8 555 9 A70209
44 304 14.8 1910 8 AF115925
45 302.2 14.7 1095 7 AF164143
```

ALIGNMENTS

```
RESULT 1
AF098484 1438 bp mRNA PRI 08-FEB-1999
LOCUS Homo sapiens napsin 1 precursor, mRNA, complete cds.
ACCESSION AF098484
VERSION AF098484.1 GI:4235424
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1438)
Koelsch,G., Wu,S., Henthorn,J., Tang,J. and Lin,X.
New human aspartic proteases napsin 1 and napsin 2: Molecular
cloning and intracellular localization of napsin 1
Unpublished
2 (bases 1 to 1438)
Koelsch,G., Wu,S., Tang,J. and Lin,X.
Direct Submission
Submitted (13-OCT-1998) Protein Studies Program, Oklahoma Medical
Research Foundation, 825 N.E. 13th Street, Oklahoma City, OK 73104,
USA
FEATURES
source Location/Qualifiers
1..1438
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
54..1316
/function="aspartic protease"
/codon_start=1
/product="napsin 1 precursor"
```

```
/protein_id="AADI3215.1"
/db_xref="GI:4235425"
/translation="MSEPPPLQPLLLLLLLLPLNVPESGATIRIPLRHVQFRRILNLL
RGWREPAELKLGAPSGEDKPIFVPLSNRYDQYFGEIGLTPQNPVFADTGSNNL
WPSRRCHFFSVPCWLFHREDPKASSSQFQNGKFAIQYGTGRVDGLSBDKLTIGGI
KGASVIFGEALWEPSSLVFAHFDGILGFLPILSVGVRPMDVLVEQGLDKVPFS
FYLNRPDEPGDELVLGGSDPAHY IPLTFVPTVYXWQHMERVKVGPGLTCAK
GCAALDITGTLITGPTTEELRALHAAIGGIPLLAGEYIILCSEIKPLIPAVSFLGGVM
FNLTAHYVYIQTRNGVRCLSGFQALDVPFPPPPFWLGLDVLGTVYVAVFDRGDMKS
SARVGLARARTRGADLWGGETAQAFPG"
sig_peptide 54..239
mat_peptide 240..1313
polyA_signal 1368..1373
BASE COUNT 302 a 422 c 406 g 308 t
ORIGIN
Query Match 65.9%; Score 1358.4; DB 88; Length 1438;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1359; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 701 gtccccagcgatgtctccacacacgcgtgctgcaacccctgctgtgctgtgctgtctgt 760
Db 44 gtcccccagcgatgtctccacacacgcgtgctgcaacccctgctgtgctgtgctgtctgt 103
QY 761 gaatgtggagcctccggggccacactgacccatccctcttcacatcagatccaactgg 820
Db 104 GAATGTGGAGCCTTCCGGGGCCACACTGATCCGATCCCTCTTCATCGAGTCCAACTGG 163
QY 821 acgcaggaccctgaacctactgaggggatggagagaaccagcagagctcccccaagtggg 880
Db 164 ACGCAGGATCTGAACCTACTAGGGGGATGAGAGAACACAGCAGAGCTCCCAAGTTGGG 223
QY 881 ggccccatccctggggacaaagccatcttcctcctcctcctcctcctcctcctcctcctc 940
Db 224 GGCCCCATCCCTTGGGGACAGCCCATCTTCGTACCTCTCTCGAAGTACAGGATGTGCA 283
QY 941 gtattttggggaaattgggctgggaacgctccacaaaacttcactgttgcctttgacac 1000
Db 284 GTATTTGGGAAATTGGGCTGGGAACGCCCTCCACAAAATTCACCTTTCCTTTGACAC 343
QY 1001 tggctctcaaatctctgtgggtcccgctcagagagatgccacttctcagtggtgcctgtg 1060
Db 344 TGCTCTCTCAATCTCTGGGTCCCGCTCCAGGAGATGCCACTTCTTCAGTGTGCCCTGCTG 403
QY 1061 gtacacacccagatttgatcccaaaagcctctagctcttccagccaatgggacaaagt 1120
Db 404 GTTACACCACCGATTGTATCCCAAGCCCTAGCTCTCTCCAGGCCAATGGGACCAAGTT 463
QY 1121 tgcattcaataygaactggggtgggtatggaatcctcctgagcgagagcaagctgactat 1180
Db 464 TGCCATTCATATGAACCTGGGGGGGTAGATGGAATCCTGAGCGAGGACAAAGCTGACTAT 523
QY 1181 tggagaatcaaggggtgatcagtgattttcggggagggtctctcctggagaccagcctggt 1240
Db 524 TGGTGAATCAAGGGTGCATCAGTGATTTCGGGGAGGCTCTCTGGGAGCCCAAGCTGCT 583
QY 1241 cttcgcttttgccattttgatgggatatattgggctcgtgtttcccatctgtctgtgaa 1300
Db 584 CTTTCGCTTTGCCCATTTTGTATGGGATATTGGCCCTCGGTTTTCCTCATCTGTCTGTGA 643
QY 1301 agagtgctggccccccgatgtagtgatgtagcagggggtctattggataagcctgtctt 1360
Db 644 AGAGTTCGGCCCCCGGATGATGTACTGTGGAGCAGGGGCTATTGGATAAGCCTGCTT 703
QY 1361 ctctctttacctcaacagggacccctgaagcctgatgagagagactgctcctgggggg 1420
Db 704 CTCCTTTTACCTCAACAGGAGCCCTGAAGAGCCTGTATGAGGAGAGAGCTGGTCTCTGGGGGG 763
QY 1421 ctgggaccggcacactacatccccccccctcactctcgtgacagtcacggtccctgccta 1480
Db 764 CTCGGACCGGACACATACATCCACCCCTACCTTCGTCGTGCCAGTCACCGTCCCTGCCTA 823
```


Db	781	ASATCCACATGGAGCGTGAAGTGGCCCGCCAGGCTGACTCTGTGCCAAGGCGTGTG	840	
Qy	1546	cggccatcctggatacagggcagctccctcclacagagacccactgaggagatccggccc	1605	
Db	841	CTGCCATCTCTGGATACAGGCACGCTCCCTCATCATCAGAGCCACTGAGGAGATCCGGGCC	900	
Qy	1606	tjcatgacagcattgggggaatccctctgctgagtgaggagatcacatcctgtctcgg	1665	
Db	901	TTCATGAGGCAATTTGGGGGAATCCCTTTGCTGGCTGGGGAGTACATCATCTGTGTCGG	960	
Qy	1666	aaatcccaaaagctcccgagctccctctctcttcttgggggggtctggtttaacctcaagg	1725	
Db	961	AAATCCCAAGCTCCCGCAGTCTCTTCTCTTGGGGGGTCTGTTTAACTCAGCG	1020	
Qy	1726	cccatgattacgtccatccagactactcgaatgagcgctccctctctgtctcgggttcc	1785	
Db	1021	CCCATGATTACGTATCCAGACTACTCGAAATGGCGTCCGCCTCTGTCTTCCGGTTTC	1080	
Qy	1786	agccctggatgtccctccgctcagggccctctctgatacctcgtgagctctcttgg	1845	
Db	1081	A3GCCCTGGAGTCCCTCCGCTCGAGGCCCTTCTGGATCCTCGGTGACGTCTCTTGG	1140	
Qy	1846	ggacgtatgtgacgctcttcgaccgctggagacatgaagagcagcgcccggtggcgctgg	1905	
Db	1141	GGACGTATGTGGCGCTCTTCGACCGCGGGACATGAAGAGCAGCGCCCGGTGGCGCTGG	1200	
Qy	1906	gagcgtcgcacitcgaggagacactcggatggagagactcgcagcgacgttcc	1965	
Db	1201	CGCGCGTCCGACATCGCGGAGCGACCTCGGATGGGAGAGACTGCCGACGCGAGTTC	1260	
Qy	1966	cgggtgacgcccaagtgaagcgcgcagcggtggtcgagaggtctcgtaccca	2025	
Db	1261	CGGGTGACGCCCAAGTGAAGCGCATCGCAGCGGTGGTGGCGAGGCTCTGTACCCA	1320	
Qy	2026	gtaaaaatccactatttccattgataaaaaa	2060	
Db	1321	GTAAAAATCCACTATTTCATTGTAATAAAAAA	1355	
RESULT	3			
E3353				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
DAVID J POWER,JOHN KEI,JEFFREY HILL,TUDY SMITH PC				
C12N15/09,A61K38/00,A61K39/395,A61K45/00,A61K48/00,PC				
C07K16/18,				
PC C12N5/10,C12P21/02,G01N33/53,G01N33/566,C12N15/00,A61K37/02,				
CC C12N5/00				
CC				
FH				
FT				
FT				
FEATURES				
source				

BASE COUNT	304	a	446	c	436	g	331	t	2	others
ORIGIN										
Query Match	65.6%; Score 1350.4; DB 10; Length 1519;									
Best Local Similarity	99.6%; Pred. No. 0;									
Matches 1354; Conservative	0;	Mismatches	6;	Indels	0;	Gaps				
Qy	701	gtccccagcagtgatctctccaccaccgctgctgcaacccctgctgctgctgctgctgctgct	760							
Db	109	GTCCCCAGCGATGTCTCCACCACCGCTGCTGCAACCCCTGCTGCTGCTGCTGCTGCTGCT	168							
Qy	761	gaatgtgagccttcgggggccacactgatcgcgatccctcttcacagtcagtcacaacctgg	820							
Db	169	GAATGTGAGCCTTCGCGGGCCACACTGATCCGCATCCCTCTTCATCGAGTCCAACTGG	228							
Qy	821	acgcagggaccctgaacctactgaggggatggagagaaaccagacagagctccccaagtggg	880							
Db	229	ACGCAGGATCCTGAACCTACTGAGGGGATGGAGAGAACCCAGCAGAGCTCCCCAAGTTGG	288							
Qy	881	ggcccatccctggggacaagcccatcttcgtacctctctcgaactacagggatgtgca	940							
Db	289	GGCCCCATCCCTCGGGACAAAGCCCATCTTCGTACTCTCTCGAACTACAGGGATGTGCA	348							
Qy	941	gtattttggggaattggcgctgggaacgctccacaaaactcactgtgtgcctttgacac	1000							
Db	349	GTATTTTGGGAAATTTGGGCTGGGAACGCTCCACAAAATTCACCTGTTCCTTTGACAC	408							
Qy	1001	tggctctccaatctctgggtcccgctccagagatgccacttctcagtgctgctgctgctg	1060							
Db	409	TGGCTCTTCCAAATCTCTGGGTCCCGTCCAGGAGATGCCACTTCTTCAGTGTGCCCTGCTG	468							
Qy	1061	gttacacacccagatttgatcccaaaagcctctagctcctccagggccaatggggaccaagt	1120							
Db	469	GTTACACCACCGATTGATTCCTCAAGCCCTCTAGCTCTTCCAGGCCAATGGGACCAAGTT	528							
Qy	1121	tggcattcaataaggaaactgggagctggatggatggatggatggatggatggatggat	1180							
Db	529	TGCCATTCAATATGGAACCTGGCGGGTAGATGGAATCTCTGAGGAGACAAAGTGACTAT	588							
Qy	1181	tggatgaatacaagggtgcatcagtgatttcggggagggctctctgggagcccgctggt	1240							
Db	589	TGGTGGAAATCAAGGTGCATCAGTATTTTCGGGGAGGCTCTGTGGAGCCAGCCTGGT	648							
Qy	1241	cttcgctttggccatcttgatgggatatgggctcgtggtttcccatctgtgtgtgga	1300							
Db	649	CTTCGTTTTTGGCCATTTTGATGGGATATTGGGCTCGGTTTCCCATTTCTGTCTGTGA	708							
Qy	1301	aggagttcgcccccggatgagtgatgagtgagcaggggctattggataaagcctgctt	1360							
Db	709	AGGAGTTTCGGCCCCCGGATGGATGACTGCTGGTGGAGCAGGGGCTATTGGATAGGCTGTCT	768							
Qy	1361	ctccttttacctcaacagggacccctgaagcctgagagcctgagagagagcctggtccggggg	1420							
Db	769	CTCCTTTTACTCAACAGGACCCCTGAAGAGCTGATGGAGGAGAGCTGCTCTGGGGGG	828							
Qy	1421	ctcggaacccggacactacatcccccctcactccttcgtgccagtcacggtccctgccta	1480							
Db	829	CTCGGAGCCGGACACTACATCCACCCCTCACCTTCGTGCCAGTCCGCTCCCTGCCTA	888							
Qy	1481	ctggcagatccacatgagcgtgtgaagtgggcccgagggctgactctctgtgccaagg	1540							
Db	889	CTGGCAGATCCACATGGAGCGTGTGAAGTGGGCCCAAGGCTGACTCTCTGTGCCAAGG	948							
Qy	1541	ctgtgctgcatcctcctggtacagggcagctccctcactcactcactcactcactcactc	1600							
Db	949	CTGTGCTGCCATCTTGGATACGGGCACGTCCTCATCACAGGACCCACTGAGGAGATCCG	1008							
Qy	1601	ggccctgcatgcagccatttgggggaatccctctgctggtggggagtagatcatcatcctgtg	1660							
Db	1009	GGCCCTGCAATGACGCCATTTGGGGAATCCCTTCTGTGGCTGGGAGTACATCATCTCTGTG	1068							

Qy	1661	ctcggaatcccaagctccccgcagtcctcttctcttcttggtggttaacct	1720
Db	1069	CTCGGAATCCCAAAGTCCCCGAGTCCTTTCCCTCTGTGGGGGGTCTGGTTAAACCT	1128
Qy	1721	cacggcccatgattactcaccagactactcgaataaggctcgctctgttgtccgg	1780
Db	1129	CACGGCCCATGATTACTGCATCAGACTACTCGAATGGCGTCCGCTCTGCTTGTCGG	1188
Qy	1781	tttcacggccctgagtccctccgctgcagggccccttctgatcctcggtagacgtctt	1840
Db	1189	TTTTCCAGGCCCTGGATGTCCTCCGCTGCAGGGCCCTTCTGGATCCTCGGTGACGCTCT	1248
Qy	1841	cttggggagctatgtgcccctcttcgaccgcggggacatgaagagcagccgccggtggg	1900
Db	1249	CTTGGGACGCTATGTGGCCCTCTTCGACC CGGGGACATCAAGACAGCGCCCGGTGGG	1308
Qy	1901	ccctggcgcgcctcgcactcgcggagcgcacctcggatggggagagaacatgcgcaggcgca	1960
Db	1309	CCTGGCGCGGCTGCGACTCGCGAGCGGACCTCGGATGGGAGAGACTGCGCAGCGCCA	1368
Qy	1961	gttccccgggtgacgcgccaaagtgaagcgcatgcgcagcggtggtcgcgagaggtctgct	2020
Db	1369	GTTCCCCGGGTGAGCCCACAGTGAAGGCATGCGCAGCGGTGCTGCGTAGTCTGCT	1428
Qy	2021	accagtaaaaatccactatttccattgaaaaaiaaaaaa	2060
Db	1429	ACCCAGTAAAAATCCACTATTTCATTGAGAAAAAACAAA	1468
RESULT 4			
E33552	E33552	1648 bp	DNA PAT 07-FEB-2001
LOCUS			
DEFINITION	ASP5.		
ACCESSION	E33552		
VERSION	E33552.1 GI:13026985		
KEYWORDS	JP 1999155583-A/1.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	David,J.P.J.J. and Jeffrey,H.T.S.S.		
JOURNAL	Asp5		
COMMENT	Patent: JP 1999155583-A 1 15-JUN-1999; SMITHLINE BEECHAM CORP OS Homo sapiens (human) PN JP 1999155583-A/1 PD 15-JUN-1999 PF 21-AUG-1998 JP 1998236037 PR 21-AUG-1997 US 60/056480,08-JUL-1998 US 09/111727 PI DAVID J POWER,JOHN KEI JEFFREY HILL,TUDY SMITH PC C12N15/09,A61K38/00,A61K39/395,A61K39/395,A61K45/00,A61K48/00,PC C07K16/18, PC C12N5/10,C12P21/02,G01N33/53,G01N33/566,C12N15/00,A61K37/02, PC C12N5/00 CC		
FEATURES	Location/Qualifiers		
source	1..1648 /organism='Homo sapiens (human)'. FH Key Location/Qualifiers FT source 1..1648 /organism='Homo sapiens (human)'.		
BASE COUNT	343 a 492 c 459 g 354 t		
ORIGIN			
Query Match	65.6%; Score 1350.4; DB 10; Length 1648;		
Best Local Similarity	99.6%; Pred. No. 0;		
Matches 1354; Conservative	0; Mismatches 6; Indels 0; Gaps		
Qy	701	gtccccagcagtgatctccacacacgctgtgctgaacccctgctgctgctgctgctgct	760

Db 1141 TATGTGGCGGTCTTCACGCGGGGACATGAAGAGCAGCGCCCGGTGGCGCTGGCGCGC 1200
Qy 1911 gctcgactcgcggagcagacctcgatgaggagagactgcagcagcagctcccccggg 1970
Db 1201 GCTCGCATCGGGAGCGGACCTCGGATGGGAGAGACTGCGCAGCGCGCAAGTTCGCCGGG 1260
Qy 1971 tgaaccccaagtgaagcgcgcagcagcgggtgggtcgcggaggtcctgtaccacagtaaa 2030
Db 1261 TGACGCCCAAGTGAAGCGCATGCGCAGCGGGTGGTGGCGGAGGTCTGCTACCCAGTAAA 1320
Qy 2031 aatccactatttccattgaaaaa 2060
Db 1321 AATCCACTATTTCACATTGAGAAAAACAAA 1350

RESULT 6
AR016588 AR016588 1299 bp DNA PAT 05-DEC-1998
LOCUS Sequence 4 from patent US 5776759.
DEFINITION AR016588
ACCESSION AR016588
VERSION AR016588.1 GI:3972865
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1299)
AUTHORS Bandman,O. and Coleman,R.
TITLE Two novel human cathepsin proteins
JOURNAL Patent: US 5776759-A 4 07-JUL-1998;
FEATURES Location/Qualifiers
1. 1299
source /organism="unknown"
BASE_COUNT 258 a 385 c 373 g 283 t
ORIGIN

Query Match 58.2%; Score 1199.2; DB 9; Length 1299;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 1286; Conservative 0; Mismatches 3; Indels 75; Gaps 1;

Qy 697 acgagctccccagcgatgtctccacacacgcgtgctcaacccctgctgtgctgctc 756
Db 3 ACGGCTCGCGAGCGATGTCTCCACCACCGCTGCTGCAACCCCTGCTGCTGCTC 62
Qy 757 tctgaatgtgagcctccggggccacactgctccgcatccctcttcagagtcacac 816
Db 63 TGCTGAATGTGGAGGCTTCCGGGGCCACACTGATCCGCATCCCTCTTCATCGAGTCCAAC 122
Qy 817 ctggagcagagacctgaacctactgagggatggagagaacacagcagagctccccaagt 876
Db 123 CTGGAGCGAGGATCTGAACCTACTGAGGGGATGGAGAGAACACAGACTCCCAAGT 182
Qy 877 tgggggccccctccccgggagacacccatcttcgtacctctctcgaactacagggatg 936
Db 183 TGGGGGCCCCATCCCTCGGGACAAGCCCATCTTCGTACCTCTCTCGAACTACAGGGATG 242
Qy 937 tgcagattttgggaattgggtgggaacgcctcccaaaacttcactgtgctttg 996
Db 243 TCAGTATTTTGGGGAATTTGGGCTGGGAACGCCCTCCACAAACTTCACATTGTGCTTTG 302
Qy 997 acactggctctcccaatctctgggtccggtccagagagatgccaactcttcagtgccct 1056
Db 303 ACACTGGCTCTCCAAATCTGGGTCCCGTCCAGGAGATGCCACTCTCTCAGTGTGCCCT 362
Qy 1057 gctggttacaccacccagttgatcccaagcctctagctcttcagagccaatgggacca 1116
Db 363 GCTGTTACACCACCGATTGATGCCAAAGCCTCTAGCTCTTCCAGGCCAATGGACCA 422
Qy 1117 agtttgccattcaatatgaaactggcggttagatgggaatcctgagcagaggaagctga 1176
Db 423 AGTTTGCCATTCAATGAACTGGCGGGTAGATGGGAATCCTGAGCGAGGACAAGCTGA 482

Qy 1177 ctattgtgggaatcaagggtgcatcagtgattttcgggagagctctctgaggagccagcc 1236
Db 483 CTATTTGGTGGATCAAGGGTGCATCAGTCAATTTTCGGGAGGCTCTCTGGGAGCCAGCC 542
Qy 1237 tggcttcgcttttcccatatttgatgggatattgggctcggttttcccatctctgtctg 1296
Db 543 TGGTCTTCGCTTTTCCCATTTTGGATGGGATATTGGGCTCGGTTTCCCATTTCTGTCTG 602
Qy 1297 tggaaaggagttcggcccccgatggatgactggtgagagcagggggtattgataagcctg 1356
Db 603 TGGAAAGGAGTTCCGGCCCCCGATGGATGACTTGGGTGGAGCAGGGGCTATTGGGAAGCCCTG 662
Qy 1357 tcttctccttttaacctcaacaggggacctgaagagcctgagagagagctggctcctgg 1416
Db 663 TCTTCTCTCTTTTACCTCAACAGGGACCTGAAGAGCCCTGATGGAGGAGAGCTGGTCTTGG 722
Qy 1417 ggggctcgagcccgccacactacataccacccctcacctctgtgcagtcacggtccctg 1476
Db 723 GGGGCTCGAGCCCGGCACACTACATCCACCCTCCCTTCGTGCGAGTCAAGGTCCCGC 782
Qy 1477 cctactgcagatccacatgagcgtgtgaagtgggccagggctgactctctgtgcca 1536
Db 783 CCTACTGGCAGATCCACATGGAGCGTGTGAAGGTGGGCCCGAGGCTGACTCTCTGTGGCA 842
Qy 1537 agggctgtgctccatcctgatacgggcacgtccctcatcacagagaccactggaggaga 1596
Db 843 AGGGCTGTGCTGCCATCTGGATACGGGCACGTCCCTCATCACAGGACCCACTGAGGAGA 902
Qy 1597 tccgggccccctgcagcagccattgggggaatcccccttgcctggctgggagtaacatcc 1656
Db 903 TCCGGGCCCCCTGCATGCAGCCATTGGGGGAATCCCTTGGCTGGGAGTAGATCATCTCC 962
Qy 1657 tgtgctcgaaaatcccaagctcccgccagctcctctctctgggggggtcgtgcttta 1716
Db 963 TGTGCTCGAAATCCCAAGCTCCCGCCAGTCTCTCTTCTTGGGGGGGTCTGGTTTA 1022
Qy 1717 acctcacggcccatgattacgtatccagactactcgaaatggcgtccgctcctctgtgt 1776
Db 1023 ACCTCAGGCCCATGATTAAGTCACTCCAGACTACTTCGAAA----- 1062
Qy 1777 ccggtttccagggccctggatgtccctccgctgcagggccctcttgatccctcgtgagc 1836
Db 1063 -----TGACG 1067
Qy 1837 tcttctgggagcgtatgtggcgtcttcgacgcggggacatgaagagcagcgcgcggg 1896
Db 1068 TCTTCTTGGGAGCGTATGTGGCCGCTTCGACCGGGGACATGAAGAGCAGCGCCCGG 1127
Qy 1897 tgggcttgccgcgcgtgcactcgcggagcggagcctcggaatggggagagactgcgcag 1956
Db 1128 TGGGCTTGGCGCGCTCGCACTCGGGAGCGGACCTCGGATGGGAGAGACTGCGCAGG 1187
Qy 1957 cgcagttccccgggtgacgcccgaagtgaagcgcatacgcagcgggtggtcgcggaggtcc 2016
Db 1188 CGCAGTTCGCCGGGTGAGCCCAAGTGAAGCGCATCGCAGCGGGTGTGTGCGGAGGTCC 1247
Qy 2017 tctaccacagtaaaatccactatttccatttcaaaaaa 2060
Db 1248 TGCTACCCAGTAAAAATCCACTATTTCATTGAAAAA 1291

RESULT 7
AR078758 AR078758 1299 bp DNA PAT 31-AUG-2000
LOCUS Sequence 4 from patent US 5965129.
DEFINITION AR078758
ACCESSION AR078758
VERSION AR078758.1 GI:10005504
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1299)
AUTHORS Bandman,O. and Coleman,R.

[illegible]


```
Qy 1491 cacatggagcgtgaaggtggccagggtgactctctgtgcaagggtgtgtgccc 1550
Db 781 CACATGGAGCGTGTGAAGTGGGTTCACGGCTGACTCTGTGCCCCAGGGCTGTGCTGCC 840
Qy 1551 atcctggatacgggcaagtcctcatcacagagaccacactgaggagatccgggcccctgcat 1610
Db 841 ATCTGTGGATACAGGCACACCTGTCTATCTGTAGAGCCCACTAGAGGATCCGGGCCCTGTGAT 900
Qy 1611 qcagccattgggggaatccccctgtgctggctgggagtagatcatcctctgtctcgaaatc 1670
Db 901 CGAGCCATTGGGGNAATCCCTTGTGCTGGGAGTACATCATCCGGTGTCTCAGAAATC 960
Qy 1671 ccaaagctcccgccagtcctctctctctctctctctctctctctctctctctctctctctct 1730
Db 961 CCAAAGCTCCCGCAGTCTCAGTTCAGGGTGACGTCCGCCCTCTGTTGTCGGGTTCGGGCC 1020
Qy 1731 gattacgtcatccagactaactcgaataggcgtccgctctctctctctctctctctctctctct 1790
Db 1021 GATTACGTCATCCAGTTCAGGGTGACGTCCGCCCTCTGTTGTCGGGTTCGGGCC 1080
Qy 1791 ctgagatgctccctccgctgcagggccctctggtatccctcgtgacgtctctctctctctctct 1850
Db 1081 TTGGACATCGCTTCGCTTCAGTACCTGTGTGGATCCTCGCGGACGTTTTCTTGGGGCG 1140
Qy 1851 tatgtgcccgtcttcgaccgcgggacatgaagagcagcgcgggtggtggtggtggtggtggtggt 1910
Db 1141 TATGTGACCGTCTTCGACCGCGGCACATGAAGAGCGCGCAGTGGGACTGGCGCGC 1200
Qy 1911 gctgcactcgcgagcgaactcggatggggagagactgcgcaggcagtcagttcccggg 1970
Db 1201 GCTGCCCTCGCGAGCGGACCTGGGAAGCGCGAGACCGCGCAGGCGCAGTACCGCGCG 1260
Qy 1971 tgagcccaagtgaagcgcgcagcgggtggtgcgaggttctctgctaccagtaaa 2030
Db 1261 TGCGCCCGAGGTGATGCGCATGCGCAGCG---GGTAGCGGAGCTAGCGCTACTCAGTAAA 1317
Qy 2031 aatccactatttccattgaaaaa 2060
Db 1318 AATCCAATATTTCCATTGAAAAA 1347

RESULT 13
A70212
LOCUS A70212 1347 bp DNA PAT 07-MAY-1999
DEFINITION Sequence 16 from Patent WO9811236.
ACCESSION A70212
VERSION A70212.1 GI:4774625
KEYWORDS .
SOURCE .
ORGANISM .
REFERENCE 1 (bases 1 to 1347)
AUTHORS Hill,J., Kay,J., and Powell,D.
TITLE ASPARTIC PROTEASE
JOURNAL Patent: WO 9811236-A 16 19-MAR-1998;
HILL JEFFREY (GB)
FEATURES
source
1..1347
/organism="unidentified"
/db_xref="taxon:3264"
BASE COUNT 262 a 386 c 388 g 306 t 5 others
ORIGIN
```

Query Match 46.18; Score 949.2; DB 9; Length 1347;
Best Local Similarity 89.4%; Pred. No. 1.5e-237;
Matches 1091; Conservative 0; Mismatches 118; Indels 11; Gaps 6;

```
Qy 698 cgcgtccccagcagtgatctccaccacccgctgctggaacccct-gctgctgctgctgctc 756
Db 44 CGGTCCCGCAGCATGTCTCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 103
Qy 757 tgcgtaatgtgagccttcgaggcgccacactgatccgcataccctctctcatagagtcacac 816
```

```
Db 104 TCCTGAATGTGGAGCTGCTGTGGGGCCACATGATCCGATCCCTCTTCTGTCAGTCCACC 163
Qy 817 ctgagcgcagagaccctgaacactactaggggatggagagaaccacagagactccccaaagt 876
Db 164 CTGGACGCAGAGACCTGAACCTACTTGAAGGGATGGGAAAAACCAAGACAGAGCTCCCCAAGT 223
Qy 877 tggggccccctccccctggggaagcccatcttcgtacctctctcgaactacagggatg 936
Db 224 TGGGGCCCCATCCCTCTGGGACAAAGCTGCTCGGTACCTCTCTCCAAATTCCTTGGATG 283
Qy 937 tgcagtatatttggggaaattgggctgggaacgcctcccaaaacttccactgttgcctttg 996
Db 284 CCCAGTATTTTGGGAAATTTGGCTTGGGAAGCGCTCCACAAAATTCCTACTGTGTGCTTTG 343
Qy 997 aactggtcctccaatctcgtggtcccggtccagagagatgcaactcttccagtgtgcct 1056
Db 344 ACATGGGTCTCTCCAATCTTGGGTCCCGTCCAGGAGATGCCACTTCTTCAAGTGTGCCCT 403
Qy 1057 gctggttacacacccgatttgatcccaagcctctagctccttcagggcccaatggagcca 1116
Db 404 GCTGTTCCACACCGCTTCAATCCCAATGCCCTCCAGCTCCTTCAAGCCCACTGGGACCA 463
Qy 1117 agttgcatatcaataggaactggcggtggttagatggaatcctgagcagagacaagctga 1176
Db 464 AGTTTGCCATTCAGTATGGAATGGCGGTGAGATGGAATCCTGAGTGAGGACAAGCTGA 523
Qy 1177 statgttggaatcaagggtgcatcagtgattttcggggagcctctctcggagaccagcc 1236
Db 524 CTATTGGTGGAAATCAAGGGTGCATCCGTGATTTTCGGGGAAGCTCTGTGGGAATCCAGCC 583
Qy 1237 tggcttcgcttttgcctcatttgatgggatattggctcgtggttttcccatctgtctg 1296
Db 584 TGGTCTTCACTGTTTCCCGCCCGATGGGATATTTGGGCTCGGTTTTCCCATTTCTGTCTG 643
Qy 1297 tggagagggttcggcccccgatggatgactggtggtgagcagggggtatggtgaagcctg 1356
Db 644 TGGAAGGAGTTTCGGCCCCCGCTGGATGTACTGTTGGAGCAAGGGCTATTGGATAAGCGCTG 703
Qy 1357 tctctcctcttaccctcaacaggaccctgaagacctgaagacctgagagagctggtccctgg 1416
Db 704 TCTTCTCTCTTTATTTCACAGGGACCTTGAAGTGGTAAATGGAGGAGAGCTGTGCTTGG 763
Qy 1417 ggggctcgagccgcacactacatccacccctcactctgctcagtcacggtccctg 1476
Db 764 GGGCTCGGACCCGCGCACACTACATCCACCCCTCAACTGCTGCGCAGTCAAGGTCCTCCG 823
Qy 1477 cctactgagatccacatgagcgtgtgaaggt-ggggccaggggtgactctctgtgccc 1535
Db 824 CCTACTGGCAGATCCACATGGAGCGTGTGAAGTGGGGCCCAAGGCTGACTCTGTGTGCC 883
Qy 1536 aaggcgtgtgctgccatcctggatacggcagctccc---tcatacagagaccactgag 1592
Db 884 AAGGG-TGTGTGCCATCTGGATACAGGACGTAACCTTCTGATCAGAGGACCCACTGAG 942
Qy 1593 gagatccgggcccctgcatgagcccatggggaaatccccctgctgctgctgctgctgctgct 1652
Db 943 GAGATCCGGGCCCTGCATGCAGCCATTTGGGGAATCCCCCTTGTGCTGGGAGTACATC 1002
Qy 1653 atcctgtgctcggaatcccaagctcccccgagtcctcctctctcttcttcttcttcttcttct 1712
Db 1003 ATCCGGTGTCTCAGAAATCCCAAGCTCCCGCAGTCTCACTCTCATTTGGGGGGGTCTGG 1062
Qy 1713 tttaacctcagggccccatgattacgtcatcagactactcgaata---tggcgctcgcctc 1769
Db 1063 TTTAATCTCAGGGGCCAGGATTTACGTCATCTCAGACTACTTCGAAAAGGGTGACGTCGCGCTC 1122
Qy 1770 tgcgtgctcgggttttcaggccctgagtgctccctcgcctgcagggccctctcttctgctcctc 1829
Db 1123 TGCTTGTCCGGCTTCAGGGCCCTTGACATCGCTCGGGCTGAAGGACCTGTCTGTGATCTCTC 1182
Qy 1830 ggtgagctctcttcttgggagcgtatgtgcccctcttcgaccgcggggagacatgaagagcagc 1889
```

Db 1183 GCGAGATTTTGGGAA--TATGTGACCGTTTGTACCGGGGACATGAAGAGCGGCGA 1240
QY 1890 gcccgggtggccctggcgcg 1909.
Db 1241 A¹CCGAGTTGACTTGGCGGG 1260
RESULT 14
AX046346 1360 bp DNA PAT 24-NOV-2000
LOCUS Sequence 13 from Patent WO0011168.
DEFINITION AX046346
ACCESSION AX046346
VERSION AX046346.1 GI:11344334
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1360)
AUTHORS Lemischka, I. and Moore, K.
TITLE Genes that regulate hematopoietic blood forming stem cells and uses thereof
JOURNAL Patent: WO 0011168-A 13 02-MAR-2000;
Princeton University (US)
FEATURES
source Location/Qualifiers
1..1360
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 293 a 385 c 355 g 327 t
ORIGIN

Query Match 38.7%; Score 797.2; DB 9; Length 1360;
Best-Local Similarity 75.6%; Pred. No. 9.4e-198;
Matches 1034; Conservative 0; Mismatches 318; Indels 16; Gaps 3;
QY 694 agaacgctcccaagcagctctccaccacgcctgcctgcgaacccctctcctcgtcgtgc 753
Db 1 AGCTAGTCCCGAGAGTGGCCACTACTGCTG-----CTGCTGCTGGCCGTCG 51
QY 754 ctctgctgaatgtggagccttcggggccacactgatccgcacccctcttcacagctcc 813
Db 52 TGTGGGGAAATTTGGAGCCTCAGAGGCCAAACTGATCCGTCCTCTTCAACGAATCC 111
QY 814 aactgacacagaccctaaactactagaggatgagagaaacacagacagactcccaca 873
Db 112 AC¹TGGACACAGAATCTTAACCCACTGAATGGATGGAGTGGGAACAGCTGGCGAGCTTTCTA 171
QY 874 agttggggggcccatccctcctgggggacaaagcccatcttcgtacactctctcgaactacagg 933
Db 172 -----GGACCTCCACCTCTGGTGGCAACCCCTCTTGTGCTCTCTCCAAAGTTATGA 225
QY 934 atjtgagtagtttggggaattggctggggaacgcctccacaaaacttcactgttgcc 993
Db 226 ACACCAGTATTTGGGAACCTATTTGGTATTTGGGAACGGCTCTCTAGAATTTACCGCTGTCT 285
QY 994 ttjacactggctcctcaactctcctgggtcccgccagagatccacactctcagtgctc 1053
Db 286 TTACACGGGTCTTCCAACTCTGGGTCCGTCACAGAGATGTCATTTCTTCAGTTGG 345
QY 1054 cc¹gctggttacaccacagattgatcccaaaagcctctagctcttcacaggccaatggga 1113
Db 346 CATGCTGGTTTACCATCGCTTTAATCCCAAGGCCTCCAGCTCCTTCAGGCCCAATGGGA 405
QY 1114 ccagtttgcattcaatgaactggcgggtgagatgaatcctcagcaggagacaagc 1173
Db 406 CCAAGTTTGCATTCAGTATFGGACCGCGGCTGAGCGGAATTCCTGACCGAGGCAATC 465
QY 1174 tgactattggtgaatcaagggtgcacatcagtgattttcggggaggctctcctgggagccca 1233
Db 466 TGA¹TATCGGGGATCCACGATGCTTTTGTGACATTTGAGAGGCTCTGTGGAGCCCA 525
QY 1234 gcctggttcctgttttgccatttgatgggatatgggctcgggttttcccaattctgt 1293

Db 526 GCCTGATCTTGTCTTAGCCACATTGATGGATCCTGGGCTCGGCTTCCCCTACTGG 585
QY 1294 ctgtggaagagttcggcccccggatggtactgtgtggagcagggctatttgataagc 1353
Db 586 CTGTGGGGGAGTTTCAGCCTCCGCTGGATGGATGGTGGAGCAAGGGCTGCTGTGAGAAC 645
QY 1354 ctgtcttctccttttactcaacagaggaacccctgaagagcctgatggaggagagctgtcc 1413
Db 646 CCGTCTTCTCTTTACTTCAACAGGATTTCTGAAGGGTCTGATGGGGGAGAGCTGTGTC 705
QY 1414 tggggggctcgaccggcgcacactacatccaccctccactctcgttcgctcgaactcagctcc 1473
Db 706 TAGGGGCTCAGACCCCGCTCACTACGTACTCCTCCCTCACTTACACAGTACCAATCC 765
QY 1474 ctgcctactggagatccacatggagcgtgtaaggtggggccagggctgactctctgt 1533
Db 766 CTGCCTACTGGCAGGTCACATGAGAGTGTGAAGGTGCGGCACAGGGCTGAGGCTGTGTG 825
QY 1534 ccaagggtcgtgctgccactcgtgatacggcgcgtccctccatcacacagagccactgagg 1593
Db 826 CCCAGGGCTCAGTGGCATCTTAGACACAGGCACATCCCTTCTCTGAATGGCAGTACTTCA 885
QY 1594 agatccggggccctgcagccatgggggaatccccctcctcgtgctgggtggggagtagatca 1653
Db 886 AGATCCGGGCTTGAATAAGCCATTGGGGGATATCCCTTCTCTGAATGGCAGTACTTCA 945
QY 1654 tctgtgctcggaatcccaaaagctcccgacgtcctctcctctcttcttctgggggggtctgt 1713
Db 946 TTCAGTGTTC¹CAAGACGCCAACGCTTCCCGCTGTCTCTCCACCTTGGTGGAGTCTGTT 1005
QY 1714 ttaacctcacggcccatgattacgtctacacactactcgaataatggctcgcctcctcgt 1773
Db 1006 TTAACCTCACAGGCCAGGACTATGTCATCAAGATTCTCAGACGATGTTGGCCTCTGCC 1065
QY 1774 tgcctgggttccaggccctggatgctcctccgcgcagggccctctcttctggtcctgggtg 1833
Db 1066 TGTGGGCTTCCAAAGCCTTGGATATCCCAAGCCTCGGGACCCCTCTCGATCTTGGGG 1125
QY 1834 acgtctcttgggagcgtatgtggcgtcttcgacgcgaggacacagaagacagagccccc 1893
Db 1126 AC¹GTCTTTTGGGCCCCCTATGTGGCTGTCTTTGACCGCTGGGACAAAACGTCGGCCCCG 1185
QY 1894 ggggtggcctggcgcgcctcgcactcgcgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1953
Db 1186 GCGTGGGACTGGCGCTGCTCAGTCTCTCAACAGACCGGGCAGAAAGAGACTACGC 1245
QY 1954 aggcgcgactccccgggtgaccccaagtga-agcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2012
Db 1246 AGCGCAGTCTTCA¹AAAGACGCCCTTGGTGGGTACAAGCTCACCGGCCACAGCAGCT 1305
QY 2013 gtctgtctaccagtaaaaaatccactatttccattgaaaaaataaaaaa 2060
Db 1306 ATGCTTCTTCCAATTAACAAACTAA¹AAAAA¹AAAAA¹AAAAA¹AAAAA¹AAAAA¹ 1353

RESULT 15

AX046376

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AX046376 1360 bp DNA PAT 24-NOV-2000
Sequence 43 from Patent WO0011168.

AX046376

AX046376.1 GI:11344363

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1360)

Lemischka, I. and Moore, K.

Genes that regulate hematopoietic blood forming stem cells and uses thereof

Patent: WO 0011168-A 43 02-MAR-2000;

Princeton University (US)

FEATURES		Location/Qualifiers	
source	1. .1360	/organism="Mus musculus"	
		/db_xref="taxon:10090"	
BASE COUNT	293 a	385 c	355 g 327 t
ORIGIN			
Query Match 38.7%; Score 797.2; DB 9; Length 1360;			
Best Local Similarity 75.6%; Pred. No. 9.4e-198;			
Matches 1034; Conservative 0; Mismatches 318; Indels 16; Gaps 3;			
Qy	694	agaacggtccccagcagatgtctccacacacgcgtgctgcaacccctgctgctgctgc	753
Db	1	AGCTTAGTCCCCAGAGATGCGCACACTACTGCTG-----CTGCTGCTGCTGCTG	51
Qy	754	ctctgctgaatgtgagacctccggggccacacatgacgcgacatccctcttcatcgagtcc	813
Db	52	TGCTGGGGAATTTGGAGCCTGAGGAGGCCAAACTGATCCGTGTCCTCTTCAACGAATCC	111
Qy	814	aacctgacgcagacacctgaacctactgagggatgagagaaacccagcagagctcccca	873
Db	112	ACCTTGGACACAGAATCTTAACCCACATGAATGATGGAAGAACAGCTGGCAGAGCTTCTA	171
Qy	874	agttggggggccccatccccctgggacaaagcccatctctctgacacctctctcgaactacagg	933
Db	172	-----GGACTCCACCTCTGCTGGCAACCCCTCTTTGTGTCCTCTCTCCAAGTTCAATGA	225
Qy	934	atgtgcagtatatttggggaaattgggtgggaacgcctcccaaaacctcaactctgtgct	993
Db	226	ACACCCAGTATTTTGGAACTATTGGTTTGGGAGCGCTCTCTCAGAAATTTACCCGTTGTCT	285
Qy	994	ttgacctggtccctccaaactctgtgggtcccgctccagagatgccactcttctcagtgtgc	1053
Db	286	TTGACACGGGTCTTCCAACTGTGGGTTCGCTCCACAGAGATGTCATTTCTCAGTTTGG	345
Qy	1054	octgctggttacacacacgatttgatcccaagcctctagctctctccaggcccaatggga	1113
Db	346	CATGCTGGTTTTCACCATCGCTTTAATCCAAAGGCTCCAGCTCCTTCAGGCCCAATGGGA	405
Qy	1114	ccaagtttgccattcaatatggaactgggggggtagatggaatccctgagcagagacaagc	1173
Db	406	CCAAATTTGGCCATTACAGTATGGGACCGGGCGGCTGAGCGGAATCCTGAGCCAGGACAAATC	465
Qy	1174	tgactattggtggaatcaaggtgcatcagtgatttccggggaggtcctctcgggagccca	1233
Db	466	TGACTATCGGGGGATCCACCATGCTTTGTGACATTTGGAGAGGCTCTGTGGAGGCCCA	525
Qy	1234	gcctggtcttcgcttttgcctatttgatgggatattgggctcggttttcccatctgt	1293
Db	526	GCCTGATCTTTGCTTTAGCCCACTTTGATGGATCCTGGGCTCGGCTTCCCACTCTGG	585
Qy	1294	ctgtggaaggagttcgcccccgatggatgtactgtgtggaagcaggggctattggataagc	1353
Db	586	CTGTGGGGAGTTTCAGCCTCCGCTGGATGGATGGTGGAGCAAGGGCTGTGGAGAAAC	645
Qy	1354	ctgtctctctctttacctcaacaggacctgaagacctgatggagagagctggtgcc	1413
Db	646	CCGCTTCTCTTTTACCTCAACAGGGATTTCTGAAGGCTCTGATGGGGGAGAGCTGGTCC	705
Qy	1414	tggggggctcgagccggcacactacatcccaacccctacacctgtgtgcccagtcacggtcc	1473
Db	706	TAGGGGGCTCAGACCCGCTCACTACGTACCTCCCTCACTTCCATACCAGTCACCATCC	765
Qy	1474	ctgctactctgcagatccacatgagcgtgtgaagggtggggcccgagggctgactctctgtg	1533
Db	766	CTGCCCTACTGGCAGGTCCACATGGAGAGTGTGAAGGTCTGAAGGGGAGAGCTGAGCCTCTGTG	825
Qy	1534	ccaagggctgtgtcccatctgatacggcactcctccatcacagaccaccactgaag	1593
Db	826	CCAGGGCTGCAGTGCCTCCTAGACACAGGCACATCCCTCATCACAGGAGCTAGTGAGG	885
Qy	1594	agatccggggccctgcatgcagccattgggggaaatccctctgtgctggtgggaglacatca	1653

Db	886	AGATCCGGGCTTGAATAAGCCATTTGGGGGATATCCCTTCTGAATGGGCAGTACTTCA	945
Qy	1654	tccgtgtcgtcggaatacccaaaagctcccccgagctctccctctcttcttgggggggtctggt	1713
Db	946	TTCAAGTGTTCNAAGACGCCAAGGCTTCCCTCTCTCTCTTCCACCTTGGTGAGTCTGCT	1005
Qy	1714	ttaaacctcaacggcccatgattacgtcatccagagactactcgaataatggcgtccgctctgct	1773
Db	1006	TTAACTCTCACAGGCCAGGACTATGTATCAAGATTTCTTTCAGAGCGATGTTGGCCTCTGCC	1065
Qy	1774	tgctcggtttccaggccctcgatgtccctccgctcgagggccctctctgataccctcggtg	1833
Db	1056	TGTTGGGCTTCCAAGCCTTGGATATCCCAAGCCTTCGGGAGCCCTCTCTGGATCCTTGGGG	1125
Qy	1834	acgtctcttggggagcgtatgtggccgtctctcgaccgcgggggacatgaagagacagcgccc	1893
Db	1126	ACGTCTTTTGGGGCCCTATGTGGCTGTCTTTGACCGTGGGCACAAAACGTCGGCCCCG	1185
Qy	1894	gggtgggacctgggcgcgctcgactcgcggagcggaacctcgatggggagagactgcgc	1953
Db	1186	GCGTGGGACTTGGCGCTGCTCAGTCTCGTTCAACAGACCGGGCAGAAAAGAGGACTACGC	1245
Qy	1954	agggcagattccccgggtgacgcccgaagtga-agcgcatgcgacggggtggtcgcggaag	2012
Db	1246	AGGCGCATTTCTTCAAAAAGACGCCCTGTAGGGTACAAAGCTCACCGGGCCACAGCAGCT	1305
Qy	2013	gtcctgctaccacagtaaaaaatccactatttccattgaaaaaataaaaaa	2060
Db	1306	ATGCTTCTTTCATTAACAAACTAAAAAATAAAAAAAAAAAAAAAAAA	1353

Search completed: November 17, 2001, 14:54:41

Job time: 7635 sec

Search completed: November 17, 2001, 14:54:41
Job time: 7635 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2001, 14:50:26 ; Search time 222.76 Seconds
(without alignments)
5806.596 Million cell updates/sec

Title: US-09-700-770-3
Perfect score: 2060
Sequence: 1 cttgagagctctcaataact.....ttccattgaaaaaataaaaaa 2060

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : -> N_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	2060	100.0	2060	21	AAZ29722 Human lung specifi
2	2060	100.0	2061	20	AAZ87152 Human protease HUP
3	1356.8	65.9	1910	19	AAV27038 Human napsin B CDN
4	1261.4	61.2	1263	21	AAZ50231 Human Aspartic Pro
5	1199.2	58.2	1298	19	AAV31665 Nucleotide sequenc
6	1151.8	55.9	1375	19	AAV28623 Human aspartic pro
7	1142	55.4	1353	19	AAV27036 Human napsin A CDN
8	1122.6	54.5	1329	21	AAZ50232 Human Aspartic Pro
9	949.2	46.1	1347	19	AAV28624 Human aspartic pro
10	797.2	38.7	1360	21	AAZ94089 Haematopoietic ste
11	797.2	38.7	1360	21	AAZ94118 Haematopoietic ste

12	716.4	34.8	1645	19	AAV27037 Human napsin A gen
13	602.4	29.2	1041	21	AAF18208 Lung cancer associ
14	397.2	19.3	481	19	AAV28634 Human aspartic pro
15	387.6	18.8	407	21	AAC00579 Human secreted pro
16	386.6	18.8	407	20	AAV40279 Human secreted pro
17	360.2	17.5	474	19	AAV28633 Human aspartic pro
18	325	15.8	408	20	AAV87699 EST clone EH106.
19	322.6	15.7	402	19	AAV28632 Human aspartic pro
20	316.6	15.4	1300	20	AAV87255 cDNA clone encodin
21	316.6	15.4	1353	21	AAA46901 cDNA encoding nove
22	316.6	15.4	2038	16	AAQ98844 Human death associ
23	316.6	15.4	2038	19	AAV60292 DNA sequence encod
24	316.6	15.4	2465	14	AAQ38758 SCD4-PCAd lysosoma
25	309.4	15.0	377	19	AAV28636 Human aspartic pro
26	307.8	14.9	2124	20	AAZ41380 Human normal uteru
27	304	14.8	555	19	AAV28635 Human aspartic pro
28	300.4	14.6	402	19	AAV28631 Human aspartic pro
29	289.2	14.0	299	20	AAV84188 DNA encoding human
30	289.2	14.0	299	21	AAC79417 5' cDNA sequence o
31	287.8	14.0	1714	21	AAC48661 Arabidopsis thalia
32	284.4	13.8	1719	21	AAC40618 Arabidopsis thalia
33	264.8	12.9	385	19	AAV28628 Human aspartic pro
34	249.2	12.1	289	21	AAZ50237 Human incyte cDNA
35	247	12.0	452	19	AAV28626 Sequence encoding
36	222.6	10.8	1459	6	AAV50180 Human aspartic pro
37	220.6	10.7	296	19	AAV28630 Human renin cDNA.
38	219.4	10.7	1358	18	AAT84705 Human incyte cDNA
39	216	10.5	227	21	AAZ50234 Human secreted pro
40	215	10.4	215	21	AAZ29306 Human gene express
41	198.8	9.7	808	20	AAZ17547 Human aspartic pro
42	191.2	9.3	500	19	AAV28625 Human incyte cDNA
43	186.8	9.1	215	21	AAZ50236 Human incyte cDNA
44	184	8.9	186	21	AAZ50233 Human incyte cDNA
45	181.4	8.8	349	19	AAV28627 Human aspartic pro

ALIGNMENTS

RESULT	1
AAZ29722	
ID	AAZ29722 standard; DNA; 2060 BP.
AC	AAZ29722;
DT	27-MAR-2000 (first entry)
DE	Human lung specific gene Lngl05.
KW	Lung Specific Gene; LSG; Lngl05; human; diagnostic marker;
OS	Homo sapiens.
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	711..1973
FT	/*tag= a
FT	/product= "LSG Lngl05 protein"
PN	WO9960160-A1.
PD	25-NOV-1999.
PF	12-MAY-1999; 99WO-US10344.
PR	21-MAY-1998; 98US-0086212.
PA	(DIAD-) DIADEXUS LLC.
PI	Yang F, Macina RA, Sun Y;
DR	WPI; 2000-116320/10.
DR	P-PSDB; AAY44457.

Db 841 tgaggagatggagagaccagcagagctcccaagtggggcccccattcctctgggaca 900
QY 901 agcccatctctacatctctcgaactacagagatgtagtgcagtaatttgggaaattggc 960
Db 901 agccatctctcgaactacagagatgtagtgcagtaatttgggaaattggc 960
QY 961 tgggaacgctccacaaactcactgttgctttgacactggctcctccaattctctgg 1020
Db 961 tgggaacgctccacaaactcactgttgctttgacactggctcctccaattctctgg 1020
QY 1021 tcccgccagagatgcaactctcagtgctccctgctggttacacacagatttgatc 1080
Db 1021 tcccgccagagatgcaactctcagtgctccctgctggttacacacagatttgatc 1080
QY 1081 ccaagcctctagctcctccagggccaatgggacaaagtttgccattcaatgaactg 1140
Db 1081 ccaagcctctagctcctccagggccaatgggacaaagtttgccattcaatgaactg 1140
QY 1141 ggcggtagatgaatcctgacgaggaacaagctactattggtgaatcaagggtgat 1200
Db 1141 ggcggtagatgaatcctgacgaggaacaagctactattggtgaatcaagggtgat 1200
QY 1201 cagtgatttccggaggctctctgggagccagcctggtctcttggccattttg 1260
Db 1201 cagtgatttccggaggctctctgggagccagcctggtctcttggccattttg 1260
QY 1261 atgggatatggcctcggtttcccatctgtctgtggaaggagttcgcccccagtg 1320
Db 1261 atgggatatggcctcggtttcccatctgtctgtggaaggagttcgcccccagtg 1320
QY 1321 atgtactggtgagcagggctattgataagcgtctctctcttcttcttcttcttct 1380
Db 1321 atgtactggtgagcagggctattgataagcgtctctctcttcttcttcttcttct 1380
QY 1381 accctgaagagcctgatggagagagctggtctctgggggctcggaacccggcacactaca 1440
Db 1381 accctgaagagcctgatggagagagctggtctctgggggctcggaacccggcacactaca 1440
QY 1441 tccacccctcacctctgacagtcacaggtccctgctactgacagatccacatgagc 1500
Db 1441 tccacccctcacctctgacagtcacaggtccctgctactgacagatccacatgagc 1500
QY 1501 gtgtgaagtgggccagggctgactctctgtgccaaggctgtgctgcccactctgata 1560
Db 1501 gtgtgaagtgggccagggctgactctctgtgccaaggctgtgctgcccactctgata 1560
QY 1561 cgggcacgtccctcacaacagaccactgaggagatccgggcccctgcatgagccattg 1620
Db 1561 cgggcacgtccctcacaacagaccactgaggagatccgggcccctgcatgagccattg 1620
QY 1621 ggggaatcccttctgctggctggagtagatcatcctctgctcgaaatcccaagctcc 1680
Db 1621 ggggaatcccttctgctggctggagtagatcatcctctgctcgaaatcccaagctcc 1680
QY 1681 ccgcagtcctctcttcttggggggtctgtgttaacctcaagcccagtcattacgtca 1740
Db 1681 ccgcagtcctctcttcttggggggtctgtgttaacctcaagcccagtcattacgtca 1740
QY 1741 tccagactactcgaatggcgtccgcctctgctgttccagggcccctggatgctc 1800
Db 1741 tccagactactcgaatggcgtccgcctctgctgttccagggcccctggatgctc 1800
QY 1801 ctcgcctcgagggccctctcggtcctggtgagctctcttctggaacatgtagccg 1860
Db 1801 ctcgcctcgagggccctctcggtcctggtgagctctcttcttgggacatgtagccg 1860
QY 1861 tcttcgaccggggacatgaagagcagcccggtgggctggcgcgctcgactc 1920
Db 1861 tcttcgaccggggacatgaagagcagcccggtgggctggcgcgctcgactc 1920
QY 1921 qcgagcagacctcgatgggagagactgcgagggcgagcttcccgggtgagcccaa 1980
Db 1921 qcgagcagacctcgatgggagagactgcgagggcgagcttcccgggtgagcccaa 1980

QY 1981 gtgaagcgcacgcgcagcgggtggtcgcgagggtcctctacccagtaaaatccactat 2040
Db 1981 gtgaagcgcacgcgcagcgggtggtcgcgagggtcctctacccagtaaaatccactat 2040
QY 2041 ttccattgaaaaaataaaaa 2060
Db 2041 ttccattgaaaaaataaaaa 2060

RESULT³

AAV27038
ID AAV27038 standard; DNA; 1910 BP.

XX AAV27038;

XX 26-OCT-1998 (first entry)

XX Human napsin B cDNA.

XX Napsin B; splicing; clone; screening; human liver cDNA library;
KW aspartic protease; N-terminal; C-terminal; genomic clone; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 526..1788

FT /*tag= a

FT /product= "napsin B protein"

XX WO9822597-A2.

XX 28-MAY-1998.

XX 20-NOV-1997; 97WO-US21684.

XX 09-MAY-1997; 97US-0046126.

XX 20-NOV-1996; 96US-0031196.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX Keolsch G, Lin X, Tang JJN;

XX WPI; 1998-312482/27.

XX P-PSDB; AAW54878.

XX New isolated aspartic protease, napsin, from human liver -
potentially useful for, e.g. diagnosis and treatment of disease

XX Claim 6; Figure 4; 24pp; English.

XX The present sequence represents the human napsin B cDNA. The N-terminus
of this cDNA was obtained by splicing together isolated napsin cDNA
clones, which had been found by screening a human liver cDNA library,
CC whereas the C-terminus was obtained by using genomic clones. Napsin B is
CC an aspartic protease which was isolated from human liver.

XX Sequence 1910 BP; 437 A; 538 C; 522 G; 412 T; 1 other;

Query Match

Best Local Similarity 65.9%; Score 1356.8; DB 19; Length 1910;
Matches 1358; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 701 gtcccagcagatgtctccaccaccgctgtgcaacccctgctgctgctgctgctgct 760

Db 516 gtcccagcagatgtctccaccaccgctgtgcaacccctgctgctgctgctgctgct 575

QY 761 gaatgtggagccttcgggggccaactgacatgcgcattcttcttcagtgacacactgg 820

Db 576 gaatgtggagccttcgggggccaactgacatgcgcattcttcttcagtgacacactgg 635

QY 821 acgcagggaccctgaacactactgaggggatggagagaaccagcagagctcccaagtgg 880

```
Db 636 acgaggaatcctgaacctactgagggtgagagaaacagcagagctccccaagtggg 695
Qy 881 ggcctccctcctgggagcaagcccatctctgtaacctctcgaactacagggatgtca 940
Db 696 ggcctccctcctgggagcaagcccatctctgtaacctctcgaactacagggatgtca 755
Qy 941 gtattttgggaaattggctgggaacgctcccaaaactcactgtgctttgacac 1000
Db 756 gtattttgggaaattggctgggaacgctcccaaaactcactgtgctttgacac 815
Qy 1001 tggctctccaatctctgggtcccgctcagagagatgccactttctcaagtgcctgtg 1060
Db 816 tggctctccaatctctgggtcccgctcagagagatgccactttctcaagtgcctgtg 875
Qy 1061 gtacacacacgatttgatcccaagcctctagctcctccagcccaatggagacaagt 1120
Db 876 gtacacacacgatttgatcccaagcctctagctcctccagcccaatggagacaagt 935
Qy 1121 tgccattcaatatggaactggcggtgtagatggaaacctgagcaggagacaagctgactat 1180
Db 936 tgccattcaatatggaactggcggtgtagatggaaacctgagcaggagacaagctgactat 995
Qy 1191 tgggtgaatacaagggtgcatcagtgattttcggggaggtcctctgaggccagcctgg 1240
Db 996 tgggtgaatacaagggtgcatcagtgattttcggggaggtcctctgaggccagcctgg 1055
Qy 1241 ctgcgttttgcctatttgatgggataattggcctcggttttccattctgtctgtgga 1300
Db 1056 ctgcgttttgcctatttgatgggataattggcctcggttttccattctgtctgtgga 1115
Qy 1301 aggagttcgcccccagtgatgtactgtgtgagcaggggctatttggaagcctgtctt 1360
Db 1116 aggagttcgcccccagtgatgtactgtgtgagcaggggctatttggaagcctgtctt 1175
Qy 1361 ctctttttacatcaacaggccctgaagacctgagtgagagagctgtgctctgggggg 1420
Db 1176 ctctttttacatcaacaggccctgaagacctgagtgagagagctgtgctctgggggg 1235
Qy 1421 ctggagcccgccacactacatccacacctcacctctgtccagtcagctccctgccta 1480
Db 1236 ctggagcccgccacactacatccacacctcacctctgtccagtcagctccctgccta 1295
Qy 1481 ctgcgagatccacatggagcgtgtgaaagtgtggccagggtgactctctgtgccaaagg 1540
Db 1296 ctgcgagatccacatggagcgtgtgaaagtgtggccagggtgactctctgtgccaaagg 1355
Qy 1541 ctgtgtgcctcctctgtgatacgggcacgtccctcctacacagaccactgaggagatccg 1600
Db 1356 ctgtgtgcctcctctgtgatacgggcacgtccctcctacacagaccactgaggagatccg 1415
Qy 1601 ggcctctgcctgcagccattgggggaatacccttgcctggctgggagtgacatcctctg 1660
Db 1416 ggcctctgcctgcagccattgggggaatacccttgcctggctgggagtgacatcctctg 1475
Qy 1661 ctgggaaatcccaagctcccgagctcctctcctcttctgggggggtctggtttaacct 1720
Db 1476 ctgggaaatcccaagctcccgagctcctctcctcttcttctgggggggtctggtttaacct 1535
Qy 1721 caagcccatgatcagtcataccagactactcgaataaggctcgcctctctgtgtccgg 1780
Db 1536 caagcccatgatcagtcataccagactactcgaataaggctcgcctctctgtgtccgg 1595
Qy 1781 ttccagggccttgatgtccctccgcctcgagggccctcttgatcctcgtgactgtgactctt 1840
Db 1596 ttccagggccttgatgtccctccgcctcgagggccctcttgatcctcgtgactgtgactctt 1655
Qy 1841 ctgggggacgtatgtggcgtcttcgaccgggggacatgaagagcagcgcccggtggg 1900
Db 1656 ctgggggacgtatgtggcgtcttcgaccgggggacatgaagagcagcgcccggtggg 1715
Qy 1901 cctggcgcgctcgcactcgcagcggaacctcggaatggggagagactgcgcaggcgca 1960
```

```
Db 1716 cctggcgcgctgcactcgcgagcggaacctcgatggggagagactgcgcaggcgca 1775
Qy 1961 gttccccgggtgacgcccgaagtgaagtcagtcgagcggtggtgcgagaggtcctgct 2020
Db 1776 gttccccgggtgacgcccgaagtgaagtcagtcgagcggtggtgcgagaggtcctgct 1835
Qy 2021 acccagtaaaatccactatttccatttccattgaaaaaataaaaaa 2060
Db 1836 acccagtaaaatccactatttccatttccattgaaaaaataaaaaa 1875

RESULT 4
AAZ50231
ID AAZ50231 standard; cDNA; 1263 BP.
XX
AC AAZ50231;
XX
DT 18-MAY-2000 (first entry)
XX
DE Human Aspartic Protease-1 (NHAP-1) encoding cDNA.
XX
KW Human; aspartic protease-1; NHAP-1; chromosome 19q13.3; antiallergic;
KW immunosuppressive; cytostatic; antiasthmatic; antinflammatory; cancer;
KW antiarteriosclerotic; antithyroid; antibacterial; neuroprotective;
KW antidiabetic; anti-HIV; osteopathic; antiarthritic; treatment; cretinism;
KW endocrinological; hypogonadism; Sheehan syndrome; diabetes insipidus;
KW hypothyroidism; adenocarcinoma; leukaemia; immunological; amyloidosis;
KW acquired immune deficiency syndrome; AIDS; Addison's disease; arthritis;
KW osteoporosis; atherosclerosis; infection; respiratory; allergy; asthma;
KW emphysema; gene therapy; diagnosis; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1263
FT
FT /*tag= a
FT /product= "Human Aspartic Protease-1 (NHAP-1)"
FT /note= "Homologous to mouse aspartic protease-like
FT protein (GI 1906810)"
FT sig_peptide 1..63
FT /*tag= b
FT mat_peptide 64..1260
FT /*tag= c
FT /product= "Mature Human Aspartic Protease-1 (NHAP-1)"
FT misc_binding 160..228
FT /*tag= d
FT /bound_moiety= "Hybridisation probe"
XX
WO200004137-A1.
XX
PD 27-JAN-2000.
XX
PF 15-JUL-1999; 99WO-US15988.
XX
PR 16-JUL-1998; 98US-0116641.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Xu H, Bruno SA, Elsenboss LA, Fogliano M, Cohan VL, Bandman O;
XX
DR WPI; 2000-182413/16.
XX
DR P-PSDB; AAY44809.
XX
PT New human aspartic protease polypeptide useful for treating and
PT detecting endocrinological disorders e.g. hypogonadism, Sheehan
PT syndrome and diabetes insipidus.
XX
PS Example 6; Fig 1; 83pp; English.
XX
CC The present cDNA sequence encodes human aspartic protease-1 (NHAP-1).
CC It is obtained from human lung cDNA library. NHAP-1 gene is located on
CC chromosome 19q13.3. NHAP has immunosuppressive, antiallergic,
CC cytostatic, antiasthmatic, antinflammatory, antiarteriosclerotic,
```


Db	782	cttactgtcagatccacatggagcgtgtgaaggtgggccaggcgtgactctctgtgcca	841
Qy	1537	agggctgtctgcataccttgatacgggcacgtccctcatcacaggaccactgaggaga	1596
Db	842	agggctgtctgcataccttgatacgggcacgtccctcatcacaggaccactgaggaga	901
Qy	1597	tcggggccctgcattcagccattgggggaatccccctgtctggctggggagtcacatccc	1656
Db	902	tcggggccctgcattcagccattgggggaatccccctgtctggctggggagtcacatccc	961
Qy	1657	tgtgtctggaaaatcccaagctcccacagtcctctcttctttttggggggctctggttta	1716
Db	962	tatctctcqaanaatcccaagctcccacagtcctctcttctttggggggctctggttta	1021

QY	I/I/	acccacagggccccatgatacacgtctatcccagacactaccgaagaaggcgccgcctcgccttgc	177
Db	1022	acctcacgggcccatgatracgtcatccagactactcgaagaa-----	1061

Qy	1177	ccggctcccaaggccctgggagctccctctgcgcgggcccctctctggagatccctccggctgacg	1896
Db	1062	-----	1066
Qy	1837	tctcttggggacgtatgtggcgtcttcgaccgcgggacatgaagacgagcccccggg	1896
Db	1067	tctcttggggacgtatgtggcgtcttcgaccgcgggacatgaagacgagcccccggg	1126
Qy	1897	tgggctggcgcgcgtcgcactcgcggagcggagcctcggatggggagagactgcgcagg	1956
Db	1127	tgggctggcgcgcgtcgcactcgcggagcggagcctcggatggggagagactgcgcagg	1186
Qy	1957	cgaagttccccgggtgaagcccaagtgaagcgcactgcgcagcgggtggtcgcggaggtcc	2016
Db	1187	cgaagttccccgggtgacgcccaagtgaagcgcactgcgcagcgggtggtcgcggaggtcc	1246
Qy	2017	tgtaccacgataaaatccactatttccattcgaataaaaaa	2060
Db	1247	tgtaccacgataaaatccactatttccattcgaataaaaaa	1290
RESULT 6			
AAV28623			
ID	AAV28623 standard; cDNA; 1375 BP.		
XX	AAV28623;		
XX	AC		
XX	AC		
DT	29-JUL-1998 (first entry)		
XX	Human aspartic protease encoding cDNA SEQ ID NO:2.		
XX	Human; aspartic protease; EST; endothelin; serum amyloid A protein;		
KW	pro-opiomelanocortin prohormone; ss.		
XX	Homo sapiens.		
OS	Homo sapiens.		
FH	Key		
FT	CDS		
FT	Location/Qualifiers		
FT	26..1339		
FT	/*tag= a		
FT	/product= "aspartic protease"		
FT	/transl_except= (pos:1322..1324,aa:Arg)		
FT	/transl_except= (pos:1325..1327,aa:Ala)		
FT	/transl_except= (pos:1328..1330,aa:Ser)		
FT	/transl_except= (pos:1331..1333,aa:Ala)		
FT	/transl_except= (pos:1334..1336,aa:Thr)		
FT	/transl_except= (pos:1337..1339,aa:Gln)		
FT	/note= "no stop codon given; the CDS is marked according		
FT	to the protein given in the specification but		
FT	could decode for its full length"		
XX	WO9811236-A1.		
XX	19-MAR-1998.		
PD	19-MAR-1998.		
XX	09-SEP-1997;		
XX	97WO-GB02426.		

RESULT	6
AAV28623	
ID	AAV28623 standard; cDNA; 1375 BP.
XX	
AC	AAV28623;
XX	
DT	29-JUL-1998 (first entry)
XX	
DE	Human aspartic protease encoding cDNA SEQ ID NO:2.
XX	
KW	Human; aspartic protease; EST; endothelin; serum amyloid A protein;
KW	pro-opiomelanocortin prohormone; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	26..1339
FT	/*tag= a
FT	/product= "aspartic protease"
FT	/transl_except= (pos:1322..1324,aa:Arg)
FT	/transl_except= (pos:1325..1327,aa:Ala)
FT	/transl_except= (pos:1328..1330,aa:Ser)
FT	/transl_except= (pos:1331..1333,aa:Ala)
FT	/transl_except= (pos:1334..1336,aa:Thr)
FT	/transl_except= (pos:1337..1339,aa:Gln)
FT	/note= "no stop codon given; the CDS is marked according to the protein given in the specification but could decode for its full length"
FT	

XX PF 09-SEP-1997; 97WO-CB02426.


```

XX OS Homo sapiens.
XX FH Key
XX FT mat_peptide
XX FT Location/Qualifiers
XX FT 31..1263
XX FT /*tag= a
XX FT /product= "Human Aspartic Protease-2 (NHAP-2)"
XX FT /note= "Homologous to mouse aspartic protease-like
XX FT protein (GI 1906810)"
XX FT 31..93
XX FT /*tag= b
XX FT mat_peptide
XX FT 94..1263
XX FT /*tag= c
XX FT /product= "Mature Human Aspartic Protease-2 (NHAP-2)"
XX FT 190..258
XX FT /*tag= d
XX FT /bound_moiety= "Hybridisation probe"
XX PN WO200004137-A1.
XX PD 27-JAN-2000.
XX XX
XX PF 15-JUL-1999; 99WO-US15988.
XX XX
XX PR 16-JUL-1998; 98US-0116641.
XX XX
XX PA (INCY-) INCYTE PHARM INC.
XX PI Xu H, Bruno SA, Elsenboss LA, Fogliano M, Cohan VL, Bandman O;
XX DR WPI; 2000-182413/16.
XX DR P-PSDB; AAY44810.
XX XX
XX PT New human aspartic protease polypeptide useful for treating and
XX PT detecting endocrinological disorders e.g. hypogonadism, Sheehan
XX PT syndrome and diabetes insipidus.
XX PS Claim 7; Fig 2; 83pp; English.
XX CC The present cDNA sequence encodes human aspartic protease-2 (NHAP-2). It
XX CC is obtained from human leucocyte cDNA library. NHAP-2 gene is located on
XX CC chromosome 19q13.3. NHAP has immunosuppressive, antiallergic,
XX CC cytostatic, antiasthmatic, antiinflammatory, antiarteriosclerotic,
XX CC antithyroid, antibacterial, neuroprotective, antidiabetic, anti-HIV,
XX CC osteopathic and antiarthritic activity. It is useful for treating and
XX CC preventing endocrinological disorders like hypogonadism, Sheehan
XX CC syndrome, diabetes insipidus, cretinism and hypothyroidism, cancers
XX CC like adenocarcinoma and leukaemia, immunological disorders like acquired
XX CC immune deficiency syndrome (AIDS), Addison's disease, amyloidosis,
XX CC arthritis, osteoporosis, atherosclerosis and microbial infections and
XX CC respiratory disorders like allergy, asthma and emphysema. NHAP
XX CC polynucleotides can be used in gene therapy and for diagnosis of
XX CC disorders associated with expression of NHAP.
XX SQ Sequence 1329 BP; 232 A; 408 C; 394 G; 295 T; 0 other;

Query Match 54.5%; Score 1122.6; DB 21; Length 1329;
Best Local Similarity 90.9%; Pred. NO. 0;
Matches 1194; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 697 acgcgtccacgcagatgtctccaccacgcgtgctgcaaacccctgcctgctgctgcctc 756
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 757 tgcgtgaatggagcttcgcggggccacactgacacactgacacactcctcttcgagtcacac 816
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 817 ctggacgcagaccctgaacctactagggatggagaaacacagacagctccccaagt 876
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 137 ctggacgcagaccctgaacctactagggggaagggaacacacagacagctccccaagt 196

```

```

QY 877 tgggggccccatccccctgggggacaagcccatcttcgtacctctctcgaactacaggatg 936
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 937 tgcagtattttggggaaattgggctgggaacgcctccacaaaacttcaactgttgcctttg 996
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 997 acactggctctccaatctctgggtccgcgtccaggagatgccacttctcaagtgcctt 1056
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 317 acactggctctccaatctctgggtccgcgtccaggagatgccacttctcaagtgcctt 376
QY 1057 gctgtttacaccaccgatttgatcccaaaacctagctctctccaggccaatggacca 1116
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1117 agtttgccattcaatatgggaactgggcgggtagatgaaactcctgagcgagacaagtga 1176
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 437 agtttgccattcagtagtgggaactgggcgggtagatgaaactcctgagtgagacaagtga 496
QY 1177 ctattgtggaatacaagggtgcacatcagtgattttcggggaggtcctctcgagagccagcc 1236
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 497 ctattgtggaatacaagggtgcacatcagtgattttcggggaggtcctctcgagagccagcc 556
QY 1237 tggctcttcgttttgcctatttgatgggatatattgggctcgtgtttcccaattctgtctg 1296
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1297 tgaagagatcggcccgcatgagatgactggtggagaggggctatttgataagcctg 1356
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 617 tggaaaggatctggcccccgtggtgactggtggagcaggggcctatttgataagcctg 676
QY 1357 tcttctctttactcaacagggaccctgaagacccctgagagcctgagtgagagagctggtctgg 1416
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 677 tcttctctttactcaacagggaccctgaagacccctgagagcctgagtgagagagctggtctgg 736
QY 1417 ggggctcggaccgcgcacactacatccaccctcactcctcgtgctcagtcacaggtccctg 1476
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 737 ggggctcagaccgcgcacactacatccaccctcactcctcgtgctcagtcacaggtccctg 796
QY 1477 cctactggcagatccacatggagcgtgtgaaagtgggccagggctgactctctgtgcca 1536
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 797 cctactggcagatccacatggagcgtgtgaaagtgggctcacggctgactctctgtgcc 856
QY 1537 agggctgtgctgcctccttggtatcaggggcacgtccctcactacacagaccactgaggaga 1596
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 857 agggctgtgctgcctccttggtatcaggggcacgtccctcactacacagaccactgaggaga 916
QY 1597 tccgggacctgcatgacgacattggggggaatcccttgcctgctggctggggagatcacatcc 1656
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 917 tccgggacctgcatgacgacattggggggaatcccttgcctgctggctggggagatcacatcc 976
QY 1657 tgtgctcggaatacccaaaagctcccgacgtctccttctcttggggggtctgattta 1716
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 977 ggtgctcagaatacccaaaagctcccgacgtctcctcactcctcattggggggtctgattta 1036
QY 1717 acctacggcccatgattacgttcacactactcgaatgcgtccgcctcctgctgtgt 1776
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1037 atctcagggccaggattacgttcacactactcgaatgcgtccgcctcctgctgtgt 1096
QY 1777 ccggtttccagaccctggatgctccctcgcctgcagggccctcttgatcctcgttgacg 1836
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1097 ccggtttcccgcccttggacatcgtctccctccagtaactgtgtggaacctcggcagc 1156
QY 1837 tctttttgggacgtatgtggccgtcttcgaccgggggacatgaaagacagcgcccggg 1896
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1157 tttctttggggcgatgtgacgtcttcgaccgggggacatgaagagcgcgacagag 1216
QY 1897 tgggacctgacgagcctcgcactcgcggagcggaacctcggatggggagagactgcgcagg 1956
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1217 tgggacctgacgagcctcgcactcgcggagcggaacctgggaagggcgagacgcgcagg 1276
QY 1957 cgcagttccccgggtgacgcggcccaagtgaagcgcatgctgcgcagcggtgtgtcgcg 2009

```


QY 1770 tctgttcggtttccaggccctgagtgctccctcgcctgcagggccctctgtgatacttc 1829
 Db 1123 tctgttcggtttccaggccctgagtgctccctcgcctgcagggccctctgtgatacttc 1182
 QY 1830 ggtgacgtcttttgggacgtatgtggcgtcttcgcagccgagacatgaagagcgc 1889
 Db 1183 ggcgaagtttttgggaa--tatgtgaccgttttgacgggggggacatgaagagcgga 1240
 QY 1890 gcccggtgggctgcgcgcg 1909
 Db 1241 acccgagttgacttgcgggg 1260

RESULT 10

AAZ94089

ID AAZ94089 standard; cDNA; 1360 BP.

XX AC

XX AAZ94089;

XX 19-JUN-2000 (first entry)

XX Haematopoietic stem cell specific nucleic acid.

XX B4-14; haematopoietic stem cell; immune system disorder;

XX leukaemia; antileukaemic; immunomodulator; therapy; mouse; ss.

XX OS Mus musculus.

XX PN W0200011168-A2.

XX XX 02-MAR-2000.

XX 20-AUG-1999; 99WO-US19052.

XX 21-AUG-1998; 98US-0138132.

XX (UYPR-) UNIV PRINCETON.

XX Lemischka I, Moore K;

XX WPI; 2000-237650/20.

XX Hematopoietic stem cell signaling proteins modulating replication and
 XX differentiation for treating immune system disorders and leukaemia -
 XX Claim 10; Page 203; 256pp; English.

CC The present sequence is that of a nucleic acid isolated from
 CC mouse primitive stem cells by methods of the invention. It is an
 CC example of claimed isolated nucleic acids (see AAZ94077-294131) that
 CC are specifically expressed in haematopoietic stem cells (HSCs) and
 CC which encode HSC-specific proteins. The HSCs are especially
 CC primitive HSCs (PHSCs) such as umbilical cord cells, bone marrow
 CC cells and foetal liver cells. The encoded proteins (see AAY79176-93)
 CC are growth factors, transcription factors, splicing factors, capping
 CC factors, transport proteins, translation factors or replication
 CC factors that modulate HSC activity, especially differentiation or
 CC replication. The invention provides a claimed method for
 CC identifying PHSC-specific nucleic acids involving: creating a PHSC
 CC cDNA library and a non-PHSC immune cell library; and subtracting
 CC the 2 libraries. Also claimed are methods: for generating a stem
 CC cell/progenitor cell from PHSCs; for identifying the presence of a
 CC PHSC in a sample; for identifying the presence in a sample of a
 CC compound that modulates HSC activity; for using such a compound to
 CC treat an immune system condition, especially leukaemia, for
 CC introducing exogenous nucleic acid into a HSC; and for ex vivo
 CC expansion of HSCs. Also claimed is a PHSC specifically expressing 1
 CC of the claimed nucleic acids, such as the present sequence.

XX SQ Sequence 1360 BP; 293 A; 385 C; 355 G; 327 T; 0 other;

Query Match

38.7%; Score 797.2; DB 21; Length 1360;

Best Local Similarity 75.6%; Pred. No. 5.1e-211;
 Matches 1034; Conservative 0; Mismatches 318; Indels 16; Gaps 3;

QY 694 agaagcggtcccccagcgatgtctccacacacccgctgtgcaacccctgctgctgtgctgc 753

Db 1 agctctagtcgccagagatgtcgccactactgtg-----ctgctgtgctgcgc 51

QY 754 ctctgctgaatgtgagaccttcggggccacacactgacgcgcatccctcttcacagatcc 813

Db 52 tgcctggggaaattggagcctgagggagggccaaactgacctgtccctcttcaagaatcc 111

QY 814 aacctggacgcagcagccctgaacctactgaggggatggagagaaaccagcagagctcccca 873

Db 112 accttgacacagaatcttaaacccactgaatggatgggaacagctggcagagcttcta 171

QY 874 agttgggggccccatccccctggggacaaagccactcttcgtacctctctcgaactacagg 933

Db 172 -----ggacctccacctctgtgtgcaacccctcttctgtcctctcctcaagtcatga 225

QY 934 atgtgcagtatatttgggaaattggcgtgggaacgcctccacaaaacttcactgttgcct 993

Db 226 acaccagtatatttggaaactattgtttgggaacgcctcctcagaatttcacogtgcct 285

QY 994 ttgacactggctctcccaatctctgggtcccgctccagagatgccactcttcagttgc 1053

Db 286 ttgacacgggtctctcccaacttctgggtcccgctccacagagatgcatttctcagttgg 345

QY 1054 cctgcgtgttacacaccccgatttgatcccaagcctctagctcctccagcccaatggga 1113

Db 346 catgctggtttccacctgcctttaaaccgaagcctccagctcctcaggcccaatggga 405

QY 1114 ccaagttgocattcaatatggaaactggcggtgagatggaatccctgagcagagacaagc 1173

Db 406 ccaagtttgccattcagtatgggacccggcgctgagcggaaacctcctgagcaggacaatc 465

QY 1174 tgactattggtgaatacaagggtgtcatcagtgattttcggggagggtctctcctggagccca 1233

Db 466 tgactatcggggggtccacagatgctttgtgacatttgagagggtctgtgaggagccca 525

QY 1234 gctggtcttcgcttttgcctatttgagggatattggcctcgttccctcctctgt 1293

Db 526 gctgactcttgccttagccactttgattgggacccctggcctcgtcctccactctgg 585

QY 1294 ctgtgaaaggagttcggcccccgcgctggtactggtgagcaggggctatttgataaagc 1353

Db 586 ctgtggcgaggagttcagctcgcgtggatgcgatgtgagcaagggtcgtcgtggagaac 645

QY 1354 ctgtctctctcttttacctcaacaggggacccctgaagagcctgatggagagctgtgctcc 1413

Db 646 cgtctctctcttttacctcaacaggggtctgaaagggtctgatgggggagagctgtgctcc 705

QY 1414 tggggggctcggaccggccacactacatccaccctcactcctcgtgcagtcacggctcc 1473

Db 706 tagggggtcagaccccgctcactacgtacctccctcactcctcaccagtcacacatcc 765

QY 1474 ctgctactggcagatccacatcggagcgtgtgaaagggtggcccgaggtgactctctgtg 1533

Db 766 ctgctactggcaggtccacatggagagtgtgaaagggtggcccgaggtgactctctgtg 825

QY 1534 ccaagggtcgtgctgccatccctggatacgggcacgtccctcctcaccagaccactgag 1593

Db 826 ccaagggtcagtgagccatccctgacacagggcacatccctcctcaccagaccactgag 885

QY 1594 agatccgggcccctgcagcagcattgggggaaatccctctgctggtggggagtagatca 1653

Db 886 agatccgggccccttgataaaagccattgggggataccctctcctgaatggcagtagtca 945

QY 1654 tccgtgctcggaaatcccaaaagctcccccagtcctcctccttcttgggggggtctggt 1713

Db 946 tccagttccaaagacgcgaacgcttcccccctgctcctcctcctcctcctcctcctcct 1005

QY 1714 ttaacctcagggcccatgatctacgtcctccagactactcgaatggcgtccgctcctcct 1773


```

RESULT 13
AAFI8208
ID AAF18208 standard; DNA; 1041 BP.
XX
XX
AC AAF18208;
XX
XX 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 227.
XX
XX Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnery;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
XX Homo sapiens.
OS
XX WO200055180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
XX WPI: 2000-587514/55.
XX P-PSDB; AAB58332.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
XX PT antigens, useful for treatment, prevention, and diagnosis of disorders
XX PT such as lung cancer
XX
XX Claim 1; Page 689; 1425pp; English.
XX
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
XX associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX associated proteins and polynucleotide sequences, their agonists, and
XX antagonists may have neuroprotective; cytostatic; cardioactive;
XX immunomodulatory; muscular active general; vulnery; gastrointestinal
XX general; nephrotropic; antiinfective; gynecological; or antibacterial
XX activity. The invention also includes antibodies specific for the
XX protein or polynucleotide sequences. The lung cancer associated
XX polynucleotide sequences may be used for detection of lung cancer,
XX chromosome identification, as chromosome markers, and for numerous other
XX diagnostic or research purposes. The proteins may be used to treat
XX disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders. The proteins may also be used in the treatment of wounds and
XX infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
XX peptide AAB58549 are used in the course of the invention for the
XX identification and characterisation of the polynucleotide and protein
XX sequences.
XX
XX Sequence 1041 BP; 227 A; 313 C; 282 G; 219 T; 0 other;

Query Match          29.2%; Score 602.4; DB 21; Length 1041;
Best Local Similarity 84.6%; Pred. No. 4.7e-157;
Matches 790; Conservative 0; Mismatches 16; Indels 128; Gaps 5;

QY 693 cagaacgcgtccacgcgagtgctccaccacgcgtgctgcaaccctgctgctgctgctg 752
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 236 cacaccacgtccacgcgagtgctccaccacgcgtgct-caaacctgctgctgctgctg 294
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 753 cctctgctgaatggtggagccttcggggccacacgtgacccgcatccctcttcacgagtc 812

```

```

Db 295 cctctgctgaatgtgga-ccttcggggccacactgacgcacccctcttcacgagtc 353
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 813 caacctggacgcagcagccctgaacctactgaggggagtgagagaaaccacgcagagctcccc 872
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 354 caacctggacgcagggatcctgaacctactgaggggagtgagagaaaccacgcagagctcccc 413
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 873 agttggggcccccattccccctggggacaagcccatctctgctacctctctcgaactacag 932
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 414 agttggggcccccattccccctggggacaagcccatctctgctacctctctcgaactacaag 473
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 933 gatgtcagttatttggggaattgggtgggaacgcctcccaaaactcactcgtgtgcc 992
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 474 g----- 474
QY 993 ttgacactgggtccctccaatctctgggtcccggtccaggagatgccactcttcagtgtg 1052
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 475 ----- 474
QY 1053 cctgtggtttacacacccagatttgatcccaaacgcctctagctcctccaggccaatggg 1112
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 475 ----gatggttacaccaccgatttgatcccaaacgcctcta--ctccttcacagccaatggg 528
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1113 accaagtgtgccaattcaaatatggaactggggcgtgtagatggaatcctgacgcagagacaag 1172
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 529 accaa-tttgccattcaatatggaactggggcgtgtagatggaatcctgacgcagagacaag 587
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1173 ctgactattggtggaatcaagggtgcatcagtgattttcggggagagctctctgaggagccc 1232
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 588 ctgactattggtggaatcaagggtgcatcagtgattttcggggagagctctctgaggagccc 647
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1233 agcctggtctctgcttttggccattttgatgggatatattgggctcggttttccattctg 1292
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 648 agcctggtctctgcttttggccattttgatgggatatattgggctcggttttccattctg 707
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1293 tctgtggaaggagttcggcccccagtgatgtagtactggtggagcaggggctattggataag 1352
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 708 tctgtggaaggagttcggcccccagtgatgtagtactggtggagcaggggctattggataag 767
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1353 cctgtctctcctttacctcaacaggggaccctgaagagcctgagagcctgagagagagctggct 1412
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 768 cctgtctctcctttacctcaacaggggaccctgaagagcctgagagcctgagagagagctggct 827
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1413 ctggggggtcctggaccgcggcacactacatccccaccctcactcctgctgcccagtcacggctc 1472
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 828 ctggggggtcctggaccgcggcacactacatccccaccctcactcctgctgcccagtcacggctc 887
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1473 cctgctactgagcagatccacatcagcgtgtgaaagtggcccagggctgactcctctgt 1532
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 888 cctgctactgagcagatccacatcagcgtgtgaaagtggcccagggctgactcctctgt 947
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1533 gccaaaggctgtgctgcccattcctgggatacgggcacgtcccctcatcacagagaccactgag 1592
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 948 gccaaaggctgtgctgcccattcctgggatacgggcacgtcccctcatcacagagaccactgag 1607
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1593 gagatccgggcccctgcatgcatgcccattgggggaa 1626
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1008 gagatccgggcccctgcatgcatgcccattgggggaa 1041
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
AAV28634
ID AAV28634 standard; cDNA; 481 BP.
XX
XX AAV28634;
XX
XX 29-JUL-1998 (first entry)
XX
DE Human aspartic protease partial cDNA EST 1320439 SEQ ID NO:12.
XX
XX Human; aspartic protease; EST; endothelin; serum amyloid A protein;
XX pro-opiomelanocortin prohormone; ss.
XX

```

OS Homo sapiens.
XX WO9811236-A1.
XX PD 19-MAR-1998.
XX PF 09-SEP-1997; 97WO-GB02426.
XX PR 11-SEP-1996; 96GB-0018966.
XX PA (SMIK) SMITHKLINE BEECHAM PLC.
XX PA (UIWA-) UNIV WALES.
XX PI Hill J, Kay J, Powell D;
XX WPI; 1998-207396/18.
XX Human aspartic protease and related DNA - which may be involved in
XX processing of endothelin and pro-opiomelanocortin pro-hormone(s)
XX Disclosure; Page 22; 32pp; English.
XX
CC The present sequence represents a partial cDNA sequence from a human
CC aspartic protease, isolated from a cDNA library of human origin. The
CC present invention describes novel human aspartic proteases and also ESTs
CC from human cDNA libraries having partial DNA sequences which encodes an
CC aspartic protease. Compounds which inhibit aspartic protease, especially
CC antibodies can be used in therapy where needed. The protease can also be
CC used in therapy where needed. Important functions of aspartic proteases
CC are the processing of endothelin and pro-opiomelanocortin prohormones.
CC They may also be involved in the processing of serum amyloid A protein.
XX
SQ Sequence 481 BP; 90 A; 132 C; 135 G; 124 T; 0 other;

Query Match 19.3%; Score 397.2; DB 19; Length 481;
Best Local Similarity 92.6%; Pred. No. 2.6e-100;
Matches 428; Conservative 0; Mismatches 33; Indels 1; Gaps 1;
QY 957 gggctgggaacccctcccaaaactcactgttgccttgactgacctggctccccaatc 1016
DB 20 gggctgggaacccctcccaaaactcactgttgccttgactgacctggctccccaatc 79
QY 1017 tgggtcccgctcagagatgccacttcttcagtgccctgctggtttacaccacgattt 1076
DB 80 tgggtcccgctcagagatgccacttcttcagtgccctgctggtttacaccacgattt 139
QY 1077 gatcccaagcctctagcttccttcaggcccaatgggaccacagtttgccattcaatgga 1136
DB 140 aatcccaatgccctcagctccttcaagccacagtgaggaccacagtttgccattcaatgga 199
QY 1137 actggcggttagatggaaatcctgagcagagacagctgactatggtggaatcaagggt 1196
DB 200 actggcggttagatggaaatcctgagcagagacagctgactatggtggaatcaagggt 259
QY 1197 gcatcagtgattttcgggagagctctctgggagccagcctgctgttcttgcctat 1256
DB 260 gcatcagtgattttcgggagagctctggtggaatccagcctgctgttcttgcctat 319
QY 1257 ttgtatggatattgggctcgtgtttcccatctctgtctgtggaaggagttcgccccc 1316
DB 320 ccgatggatattgggctcgtgtttcccatctctgtctgtggaaggagttcgccccc 379
QY 1317 atggatgactgttgagcagggctattgataagcctgtcttcttcttacctcaac 1376
DB 380 ctggatgactgttgagcagggctattgataagcctgtcttcttcttacctcaac 439
QY 1377 agggacccttg-aagagccttgatggaggagagctggtctctggg 1417
DB 440 agggacccttgatggaggagagctggtctctggg 481

RESULT 15

AAC00579
XX ID AAC00579 standard; cDNA; 407 BP.
XX AC AAC00579;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 577.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PA (GEST) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-500381/45.
XX DR P-PSDB; AAG00573.
XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 577; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 407 BP; 79 A; 137 C; 97 G; 92 T; 2 other;

Query Match 18.8%; Score 387.6; DB 21; Length 407;
Best Local Similarity 99.0%; Pred. No. 1.1e-97;
Matches 398; Conservative 2; Mismatches 1; Indels 1; Gaps 1;
QY 701 gtccccagcgatgtctccaccaccgctgtgcaacccctgctgctgtcctctgtctgt 760
DB 7 gtccccagcgatgtctccaccaccgctgtgcaacccctgctgctgtcctctgtctgt 66
QY 761 gaatgtgagccttcggggccacactgacgacacccctcttcacgagtcacacctgg 820
DB 67 gaatgtgagccttcggggccacactgacgacacccctcttcacgagtcacacctgg 126
QY 821 acgcaggaccctgaacctactgaggggagagaaacacagcagagctcccaagtggg 880
DB 127 acgcaggacccctgaacctactgaggggagagaaacacagcagagctcccaagtggg 186
QY 881 ggccccatccccctggggaagaagcccatctctctacatctctcgaactacaggatgca 940
DB 187 ggccccatccccctggggaagaagcccatctctctacatctctcgaactacaggatgca 246

